



UNIVERSIDADE NOVA DE LISBOA

**DESENVOLVIMENTO DE UM ENSAIO DE HIBRIDAÇÃO
MÚLTIPLA (*MHA – MULTIPLE REGION HYBRIDIZATION
ASSAY*) PARA IDENTIFICAÇÃO PRESUMÍVEL DE VÍRUS
DA IMUNODEFICIÊNCIA HUMANA DO TIPO 1 (HIV-1)
DOS SUBTIPOS B, G E DE FORMAS RECOMBINANTES
CRF14_BG E CRF02_AG EM PORTUGAL**

Ferdinando Bernardino De Freitas

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE EM MICROBIOLOGIA MÉDICA

ABRIL 2010



UNIVERSIDADE NOVA DE LISBOA

**DESENVOLVIMENTO DE UM ENSAIO DE HIBRIDAÇÃO
MÚLTIPLA (*MHA – MULTIPLE REGION HYBRIDIZATION
ASSAY*) PARA IDENTIFICAÇÃO PRESUMÍVEL DE VÍRUS
DA IMUNODEFICIÊNCIA HUMANA DO TIPO 1 (HIV-1)
DOS SUBTIPOS B, G E DE FORMAS RECOMBINANTES
CRF14_BG E CRF02_AG CIRCULANTES EM PORTUGAL**

Ferdinando Bernardino De Freitas

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE EM MICROBIOLOGIA MÉDICA

Orientador: Professor Doutor Ricardo Parreira

Unidade de Virologia

Instituto de Higiene e Medicina Tropical

ABRIL 2010

Apresentação em Congresso, (Formato *Poster*): (ver Anexo 3)

F.B.Freitas, A. Esteves, J. Piedade and R. Parreira. 2009. Development of a multi-region hybridization assay (MHA_{BG0214}) for the identification of subtypes B, G and recombinant forms CRF02_AG and CRF14_BG of the human immunodeficiency virus type 1 (HIV-1) circulating in Portugal.

Book of Abstracts of Microbiotec09, Poster Session: S3-Clinical Microbiology and Epidemiology, 28-30 Novembro, 2009, Vilamoura, Portugal.

Microbiotec09, 2009, Centro de Congressos do Tivoli Marina Hotel, Vilamoura, Portugal

AGRADECIMENTOS

Ao meu orientador, Professor Doutor Ricardo Parreira, pela transmissão de conhecimentos, confiança, amizade, constante disponibilidade e pelos bons momentos que partilhamos ao longo da realização deste trabalho.

À Professora Doutora Aida Esteves pelo incentivo e primeiro contacto com os temas no âmbito da tese.

A todos os elementos da Unidade de Virologia, nomeadamente ao Professor Doutor João Piedade, Dr.^a Susana Valente, Teresa Venenno, Fernanda e Cidália, pelos conselhos, bons momentos e ajuda no laboratório.

Aos meus colegas de mestrado, Dr.^a Cristina Branco, Dr.^a Raquel, Dr.^a Sandra Castro, Dr.^a Rita, Eng.^a Susana Costa, Dr.^a Filipa Matos, Dr.^a Margarida, Dr. João Almeida, Dr.^a Diana Machado, Dr. Jorge Ramos pela ajuda, momentos de descontração e alegria.

Aos investigadores, colegas e amigos da Unidade de Entomologia Médica, do CMDT/UEI Malária e da Unidade de Micobactérias, por todo o apoio e amizade durante a realização deste projecto, nomeadamente o, Dr. João Rodrigues, Dr.^a Sofia Borges, Dr.^a Diara Lopes, Eng.^o. José Vicente, Eng.^a Paula Figueiredo, Dr.^a Louise Rodrigues, Dr.^a Filipa Ferreira, Doutora Teresa Novo, Professor Doutor Paulo Almeida, Professor Doutor Pedro Cravo, Dr.^a Sofia Costa, Doutora Gabriella Spengler, Dr.^a Liliana Rodrigues, Dr. Nuno Martins, Dr.^a Ana Côrte-Real.

Aos meus colegas e amigos da UMa, Miguel Nóbrega, Ana Mota, Cátia Vieira, Odília Jardim, um obrigado pelos momentos partilhados, pelos sorrisos, pela ajuda e compreensão nos momentos difíceis.

Aos meus amigos, Eurico Santos, Nuno Vieira, Carlos Figueira, Marco Freitas e Xávier pelos momentos de alegria, pela amizade, incentivo e sobretudo pela compreensão em relação aos meus compromissos académicos.

À Ana Rosário, pelo apoio, exemplo, pela amizade e pelas importantes discussões, que sem duvida contribuíram para a minha formação e motivação para o trabalho.

À minha família, em especial ao meu irmão Carlos e aos meus avós Nuno e Filomena, um profundo e sincero obrigado, porque sem o seu apoio, a sua palavra amiga e sobretudo confiança não teria sido possível chegar até aqui.

À Cristina pela grande ajuda na realização deste trabalho, incentivo e pelo que representa para mim.

Por fim queria agradecer a todos os que de alguma forma contribuíram e continuam a contribuir para o meu crescimento profissional e pessoal, aos que me inculcaram valores e me ensinaram a respeitar e a ser respeitado, Obrigado...

Dedico esta tese à minha Mãe...

RESUMO

A maioria dos métodos utilizados na caracterização genética do HIV-1 baseia-se na análise de regiões específicas do genoma viral, fornecendo informação parcial sobre o mesmo e, por consequência, revelando-se inadequados para a identificação de vírus recombinantes. O único método que permite uma caracterização integral do genoma viral passa pela sua sequenciação completa. No entanto, este é um método dispendioso, laborioso e de difícil implementação quando se pretende a análise de elevados números de amostras. Como alternativa a este último, o conjunto de métodos genericamente designados de MHA (*Multiple Region Hybridization Assay*) baseiam-se na amplificação, por PCR em tempo-real, de várias regiões ao longo do genoma viral e na sua caracterização com sondas específicas (*TaqMan*). Tendo este modelo por base, o objectivo deste estudo foi o desenvolvimento de um ensaio de hibridação múltipla (MHA_{BG0214}) passível de ser aplicado ao estudo de um elevado número de amostras. Este método foi desenvolvido tendo como objectivo a genotipagem as estirpes circulantes dominantes na epidemia Portuguesa, nomeadamente os subtipos B, G e formas genéticas recombinantes CRF02_AG e CRF14_BG.

Com base em alinhamentos de sequências de referência de genoma completo, delinearam-se *primers* universais e subtipo-específicos para a amplificação de diversas regiões codificantes distribuídas ao longo do genoma do HIV-1 (Gag, Protease, Transcriptase Reversa, Integrase, Rev, Gp120 e Gp41). A optimização foi efectuada, inicialmente, para um conjunto de amostras de referência e seguidamente avaliada num conjunto de 50 amostras clínicas. O MHA_{BG0214} foi implementado numa estratégia de PCR em tempo-real, numa detecção dependente de *SYBR[®] Green I* para todas as regiões ou, como alternativa, usando sondas *TaqMan* (Gp41). Apresentamos ainda uma estratégia em que a análise de resultados se baseia, simplesmente, numa abordagem usando PCR/gel de agarose convencional. Estas abordagens constituem ferramentas úteis na identificação das estirpes de HIV-1 em Portugal.

ABSTRACT

Since most methods used for the genetic characterization of HIV-1 are based on the analysis of singular regions of the viral genome, they only provide a fragmented view of the later, and often fail to identify recombinant viruses. The most efficient method for HIV-1 genetic characterization involves full-genome sequencing, but the associated costs and low throughput preclude it from being routinely used for the analysis of large numbers of viral strains. One of the alternatives for a consistent genetic analysis of large numbers of viral strains are the methods generally known as Multi Region Hybridization Assays (MHA). MHA relies on the amplification by real-time PCR of several regions scattered along the HIV-1 genome, and their characterization subtype-specific TaqMan probes. In this context, the aim of our study was the development of a new multi-region hybridization assay (MHA_{BG0214}) for genotyping of the major HIV-1 forms circulating in Portugal, subtypes B, G and CRF02_AG and CRF14_BG circulating recombinant viruses.

Based on full alignments of representative HIV-1 reference sequences, we designed universal and subtype-specific primers/probes for the amplification of several different coding regions of the viral genome (Gag, Protease, Reverse Transcriptase, Integrase, Rev, Gp120 and Gp41). Optimization of reaction conditions, established using 7 HIV-1 references, served as a starting point for the analysis of 45 HIV-1 strains circulating in Portugal. MHA_{BG0214} was implemented using a real-time PCR-based approach, with detection dependent on the use of either *SYBR*[®] *Green* I (all regions) or a TaqMan probe (Gp41). Alternatively, a technically less demanding strategy based on conventional PCR and agarose gel analysis of reaction products was also developed. Regardless of the strategies used and the results obtained, these approaches are useful tools to identify strains of HIV-1 in Portugal.

ÍNDICE GERAL

AGRADECIMENTOS.....	ii
RESUMO.....	iv
ABSTRACT.....	v
ÍNDICE GERAL.....	vi
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE TABELAS.....	xii
ABREVIATURAS.....	xiii
1. Introdução	2
1.1 A síndrome da imunodeficiência adquirida	2
1.2 O vírus da imunodeficiência humana do tipo 1	5
1.2.1 Estrutura morfológica do virião	5
1.2.2 Organização do genoma viral e funções das proteínas virais.....	6
1.2.3 Ciclo replicativo do HIV-1	9
1.3 Origem do HIV-1	11
1.4 Diversidade genética.....	12
1.4.1 Mecanismos de variação genética.....	12
1.4.2 Classificação das estirpes de HIV-1.....	13
1.4.3 Distribuição geográfica global das estirpes HIV-1 do grupo M	15
1.4.4 Diversidade genética das estirpes de HIV-1 circulantes em Portugal	17
1.5 Impacto da diversidade genética na epidemia do HIV-1	18
1.5.1 Patogenicidade, transmissibilidade e progressão para a doença	18
1.5.2 Resistência aos fármacos com actividade antirretroviral	19
1.5.3 Desenvolvimento de vacinas.....	20
1.5.4 Impacto da diversidade genética do HIV na sensibilidade dos métodos de diagnóstico e de monitorização da infecção (determinação de carga viral)	22
1.6 Métodos de genotipagem de HIV-1	23

1.6.1 Genotipagem de HIV-1 baseados na análise da mobilidade diferencial de homo e heteroduplexes.	24
1.6.2 Métodos de genotipagem de estirpes HIV-1 recorrendo à incorporação de moléculas fluorescentes nas cadeias de DNA	27
1.6.3 Genotipagem de HIV-1 por sequenciação do genoma viral.....	28
1.6.4 Genotipagem de HIV-1 através da utilização de métodos de hibridação múltipla	30
1.7 Notas finais e objectivos	36
2. Material e Métodos	38
2.1 Amostras de material biológico utilizadas neste estudo.....	38
2.1.1 Caracterização de amostras de referência	38
2.1.2 Caracterização de amostras biológicas de origem clínica.....	39
2.2 Preparação de amostras biológicas	40
2.2.1 Preparação de amostras de referência	40
2.2.1.1 Preparação de células competentes de <i>Escherichia coli</i> e sua transformação por choque térmico	41
2.2.1.2 Isolamento de plasma a partir de amostras de sangue periférico total	42
2.2.1.3 Preparação das amostras de sangue imobilizado em cartões FTA™ (Whatman/GE Healthcare) .	42
2.3 Extracção de ácidos nucleicos	43
2.3.1 Extracção de ácidos nucleicos de amostras de referência	43
2.3.2 Extracção de RNA total a partir de plasma humano congelado a -80°C.....	43
2.4 Quantificação espectrofotométrica de amostras de DNA	44
2.5 Síntese, por reacção de transcrição reversa, de DNA complementar (cDNA) a partir de RNA total ..	45
2.6 Construção de alinhamentos múltiplo de sequências de HIV-1 para delineamento de <i>primers</i> /sondas	46
2.7 Delineamento de <i>primers</i> para PCR e de sondas para PCR em tempo-real	46
2.8 Amplificações de fragmentos de DNA por reacção em cadeia da Polimerase (PCR)	47
2.8.1 Verificação, por PCR, da integridade estrutural das amostras de referência (clones plasmídicos de genoma total ou parcial).....	47
2.8.2 Amplificação de fragmentos de DNA (universais e subtipo-específicos) para as várias regiões em estudo por PCR: amostras de referência.....	49
2.8.3 Amplificação de sequências virais por PCR a partir de amostras biológicas.....	51
2.8.4 Amplificação dos fragmentos de DNA por PCR em tempo-real: ensaio de competição de matrizes (amostras de referência)	52

2.8.5 Detecção de amplificação específica de sequências de HIV-1 (amostras biológicas), por PCR em tempo-real realizado na presença de <i>SYBR[®] Green I</i>	55
2.8.6 Amplificação específica de sequências virais (Gp41) de HIV-1 (amostras clínicas) por PCR em tempo-real usando sondas de hidrólise (<i>TaqMan</i>).....	56
2.9 Amplificação, purificação e sequenciação de produtos de PCR.....	57
2.10 Análise filogenética das sequências nucleotídicas	58
2.11 Análise de recombinantes	59
3. Resultados	61
3.1 Preparação das amostras de referência: extração e quantificação de ácidos nucleicos	61
3.2 Delineamento de <i>primers</i> /sondas para PCR a partir de alinhamentos múltiplos de sequências de HIV-1	62
3.3 Amplificações de fragmentos de DNA por PCR usando amostras de referência	66
3.3.1 Verificação, por PCR, da integridade estrutural das amostras de referência utilizadas	66
3.3.2 Amplificação de fragmentos de DNA universais externos e internos para as várias regiões em estudo	67
3.3.3 Amplificação específicas (subtipo/CRF) de diferentes regiões alvo distribuídas ao longo do genoma do HIV-1	70
3.3.4 Amplificação de sequências genômicas de HIV-1 por PCR em tempo-real: ensaio de competição de matrizes	74
3.4 Amplificação por PCR de sequências virais alvo a partir de amostras clínicas.....	76
3.4.1 Obtenção de fragmentos universais externos e internos	76
3.4.2 Amplificação por PCR de sequências virais específicas.....	78
3.4.3 Utilização do PCR em tempo-real para amplificação específica de sequências de HIV-1 e detecção, em simultâneo, dos produtos de amplificação.....	82
3.4.4 Detecção de amplificação específica de sequências virais (Gp41) de HIV-1 por PCR em tempo-real usando sondas de hidrólise (<i>TaqMan</i>).....	85
3.5 Confirmação da identidade genética das estirpes virais presentes em amostras clínicas: caracterização de cada região do genoma por análise filogenética.....	87
3.5.1 Amplificação, purificação e sequenciação de produtos de PCR.....	87
3.5.2 Análise filogenética das sequências virais	89
3.5.3 Análise comparada dos resultados de genotipagem baseados em PCR com <i>primers</i> específicos e da análise filogenética das sequências estudadas.....	92
3.5.4 Avaliação, por <i>bootscanning</i> , da natureza recombinante de algumas das sequências de HIV-1 analisadas	97

4. Discussão e Conclusões	103
5. Referências bibliográficas.....	133
Anexo.....	150

ÍNDICE DE FIGURAS

Figura 1.1 – Representação gráfica da história natural da infecção pelo HIV.	3
Figura 1.2 – Mapa representativo da variação na prevalência da infecção pelo HIV (do tipo 1 e 2) a nível Mundial	4
Figura 1.3 – Esquema representativo da estrutura do virião maduro de HIV-1.....	6
Figura 1.4 – Representação esquemática do genoma da estirpe HXB2 de HIV-1 (estirpe de referência). ..	7
Figura 1.5 – Representação esquemática simplificada dos principais passos do ciclo replicativo do HIV-1	10
Figura 1.6 – Diversidade genética do HIV-1	14
Figura 1.7 – Distribuição geográfica dos vários subtipos predominantes e formas recombinantes circulantes (CRFs) de HIV-1 do grupo M	16
Figura 1.8 – Representação esquemática dos genomas das estirpes de HIV-1 mais frequentemente encontradas em Portugal.....	17
Figura 1.9 – Esquema representativo dos pontos-chave de um ensaio de HMA para subtipagem do HIV-1.	25
Figura 1.10 – Representação esquemática de um protocolo experimental para a amplificação do genoma completo de HIV-1.....	28
Figura 1.11 – Esquema de árvores filogenéticas para a análise filogenética do HIV-1	30
Figura 1.12 – Esquema representativo do protocolo experimental utilizado num ensaio de MHA.	32
Figura 1.13 – Representação esquemática genérica do sistema envolvido na detecção de sinal fluorescente emitido como resultado da utilização de sondas tipo <i>TaqMan</i> numa plataforma de PCR em tempo-real	33
Figura 1.14 – Esquema representativo da construção de sondas específicas a implementar em ensaios de MHA.....	34
Figura 2.1 – Esquema representativo da estratégia amplificação utilizada na amplificação universal e específica dos fragmentos de DNA para as várias regiões do genoma do HIV-1.	49
Figura 2.2 – Representação esquemática dos principais passos para a amplificação por PCR em tempo-real na presença de <i>SYBR® Green I</i>	53
Figura 3.1 – Alinhamento múltiplo de sequências de estirpes de referência de HIV-1.	63
Figura 3.2 – Esquema representativo das várias regiões analisadas ao longo do genoma.	65
Figura 3.3 – Visualização dos produtos de PCR resultantes da amplificação da região RT para as amostras de referência (avaliação da integridade do genoma).	67
Figura 3.4 – Exemplo do perfil de visualização dos produtos amplificados para os fragmentos externos e internos (região da Pr)	69

Figura 3.5 – Exemplo do perfil de amplificação específica obtido para duas amostras de referência (subtipos B e G).....	71
Figura 3.6 – Exemplo de um ensaio de competição por PCR em tempo- real entre duas matrizes de referência (região de Gp41).....	76
Figura 3.7 – Observação em gel de agarose dos produtos de amplificação para as amostras clínicas (fragmentos universais externos e internos).	77
Figura 3.8 – Amplificação dos fragmentos específicos referentes às amostras clínicas (região PR).....	80
Figura 3.9 – Ensaio de amplificação específica por PCR em tempo-real. (Regiões Pr e Gp120).....	83
Figura 3.10 – Ensaio de amplificação específica por PCR em tempo-real. (Regiões RT e Gp41).....	85
Figura 3.11 – Ensaio de amplificação específica por PCR em tempo-real com sonda de hidrólise (<i>TaqMan</i>) para a região Gp41	87
Figura 3.12 – Relações filogenéticas entre estirpes de HIV-1 baseadas na análise das sequências nucleotídicas da região codificante da protease viral (amostras clínicas)	91
Figura 3.13 – Análise de <i>bootscanning</i> de sequências virais (região Gag) (amostras clínicas).....	99
Figura 3.14 – Análise de <i>bootscanning</i> de sequências virais (região Pr) (amostras clínicas).	100
Figura 3.15 – Análise de <i>bootscanning</i> de sequências virais (região RT) (amostras clínicas).	101
Figura 3.16 – Análise de <i>bootscanning</i> de sequências virais (região Gp120) (amostras clínicas).....	101

ÍNDICE DE TABELAS

Tabela 2.1 - Condições utilizadas para as reacções de transcrição reversa a partir de RNA total.	45
Tabela 2.2 - Condições utilizadas para a amplificação por PCR dos fragmentos de DNA para as cinco regiões do genoma das estirpes virais de referência	48
Tabela 2.3 - Condições utilizadas para amplificação, por PCR, dos fragmentos externos e internos para as várias regiões do genoma do HIV-1 em estudo.	51
Tabela 2.4 - Condições utilizadas para amplificação dos fragmentos específicos para as várias regiões do genoma do HIV-1 em estudo por reacção de PCR.	51
Tabela 2.5 - Condições utilizadas para amplificação de secções do genoma de HIV-1 por PCR em tempo- real.	55
Tabela 2.6 - Condições utilizadas para a detecção específica de parte da sequência codificante de Gp41 por PCR em tempo-real usando sondas <i>TaqMan</i>	57
Tabela 3.1 – Concentração do DNA correspondente aos fragmentos externos para as amostras de referência	62
Tabela 3.3 – Resumo dos resultados obtidos para a amplificação dos fragmentos externos e internos para as amostras de referências por todas as regiões analisadas do genoma	69
Tabela 3.4 – Temperaturas e tamanho do fragmento esperado de acordo com o <i>primers</i> subtipo/CRF específicos utilizados e respectiva região do genoma.	74
Tabela 3.5 – Temperaturas de hibridação utilizadas na amplificação por PCR de sequências virais específicas, re-otimizadas com base nos resultados de amplificação obtidos para amostras clínicas.	79
Tabela 3.6 – Resumo dos resultados obtidos para as amplificações específicas a partir das amostras clínicas ao longo das regiões analisadas	81
Tabela 3.7 – Resumo dos resultados obtidos para as amplificações específicas por PCR em tempo-real a partir das amostras clínicas ao longo das regiões analisadas	84
Tabela 3.8 – Dimensões dos fragmentos esperados para cada uma das regiões analisadas (amostras para sequenciação e análise filogenética).	88

ABREVIATURAS

A,C,G,T	Bases orgânicas constituintes dos nucleótidos (a denina, c itosina, g uanina e t imina, respectivamente)
AS	Hospital Professor Doutor Fernando da Fonseca (A madora- S intra)
BSL	<i>BioSafety Level</i>
BST-2	Do inglês <i>Bone Marrow Stromal cell antigen 2</i>
°C	Graus Celsius
CD4⁺	Do inglês <i>Cluster of Differentiation 4</i>
CD8⁺	Do inglês <i>Cluster of Differentiation 8</i>
CDC	<i>Centers for Disease Control and Prevention</i>
cDNA	DNA complementar
cpz	Chimpanzés
CRF	Forma recombinante circulante, do inglês <i>Circulating Recombinant Form</i>
DEPC	Dietil pirocarbonato
DNA	Ácido desoxirribonucleico, do inglês <i>Deoxyribonucleic Acid</i>
dNTP	Desoxirribonucleótidos 5'-trifosfatados
env	Glicoproteína do invólucro, do inglês, <i>Envelope glycoprotein</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Etilenodiaminotetracetato
ELISA	Do inglês, <i>Enzyme Linked Immuno Sorbent Assay</i>
g	grama
<i>g</i>	Aceleração da gravidade
<i>gag</i>	Do inglês, <i>group-specific antigen</i>
h	Hora
HAART	<i>Highly Active AntiRetroviral Treatment</i>
HIV-1	Vírus da imunodeficiência humana tipo 1, do inglês <i>Human Immunodeficiency Virus type 1</i>
HIV-2	Vírus da imunodeficiência humana tipo 2, do inglês <i>Human Immunodeficiency Virus type 2</i>
HMA	Do inglês <i>Heteroduplex Mobility Assay</i>

HTA	Do inglês <i>Heteroduplex Tracking Assay</i>
INSA	Instituto Nacional de Saúde Dr. Ricardo Jorge
IPTG	Isopropil- β -D-tiogalactopiranosídeo
LTR	Repetição terminal longa, do inglês <i>Long Terminal Repeat</i>
K2P	Kimura a dois parâmetros
KCL	Cloreto de potássio
kDa	Kilo(Daltons) = 10^3 Daltons
k(p)b	Kilo(pares) de bases = 10^3 (pares) de bases
LB	Meio de Luria-Bertani
M	Molaridade
MgCl₂	Cloreto de magnésio
m	Mili, prefixo que designa 10^{-3}
M	Molar
MHA	Do inglês, <i>Multiregion Hybridization Assays</i>
min	Minuto
ml	Mililitro
mM	Milimolar
mRNA	Ácido ribonucleico mensageiro, do inglês <i>Messenger Ribonucleic Acid</i>
ng	Nanograma
(NH₄)₂SO₄	Sulfato de amónio
NIAID	<i>National Institute of Allergy and Infectious Diseases</i>
NJ	<i>Neighbor-Joining</i>
nm	Nanómetro = 10^{-9} metros
nM	Nanomolar
NNRTI	Do inglês <i>Non-Nucleoside Reverse Transcriptase Inhibitors</i>
NRTI	Do inglês, <i>Nucleoside Reverse Transcriptase Inhibitors</i>
nt	Nucleótido
pb	Pares de bases
PBMC	Células mononucleadas do sangue periférico, do inglês <i>Peripheral Blood Mononuclear Cells</i>
PCR	Reacção de polimerização em cadeia, do inglês <i>Polymerase Chain Reaction</i>

pH	Simétrico do logaritmo decimal da concentração hidrogeniónica de uma solução
pol	Do inglês, <i>viral polymerase</i>
pmol	Picomole = 10^{-12} moles
p/v	Massa de soluto por volume de solução
rHMA	Do inglês <i>reverse HMA</i>
RNA	Ácido ribonucleico, do inglês <i>Ribonucleic Acid</i>
RNase	Ribonuclease
RT	Do inglês <i>Reverse Transcription</i>
RT-PCR	Polimerização em cadeia da polimerase antecedita de transcrição reversa
seg	Segundos
SIDA	Síndrome da I munodeficiência A dquirida
SIV	Vírus da imunodeficiência dos Símios, do inglês <i>Simian immunodeficiency viruses</i> .
TA	Centro de Atendimento a Toxicodependentes das T aipas em Lisboa
TAE	Tampão Tris-acetato EDTA
TE	Tampão Tris-EDTA
TM	Temperatura de desnaturação (do inglês T emperature M elting)
Tris	Tris(hidroximetil)aminometano
Tris-HCL	Tris Hidrocloreto
TSS	<i>Transformation and Storage Solution</i>
UDI	Utilizadores de D rogas I njectáveis
URF	Do inglês <i>Unique Recombinante Forms</i>
UV	Ultravioleta
μ	Micro, prefixo que designa 10^{-6}
μg	Microgramas
μl	Microlitros
μM	Micromolar

1. Introdução

1. Introdução

Os primeiros casos de uma doença complexa que viria a ser denominada de Síndrome de **Imunodeficiência Adquirida** (SIDA) foram documentados, pela primeira vez, em 1981, na sequência do diagnóstico de um conjunto de infecções devidas a *Pneumocystis jirovecii* (anteriormente designado por *Pneumocystis carinii*), citomegalovírus e *Toxoplasma gondii*, e à detecção de linfomas e sarcoma de Kaposi em indivíduos jovens, sem quaisquer indícios prévios de imunodeficiência (12, 39). Alguns anos mais tarde (1983-1984) foi isolado o agente etiológico responsável pela SIDA, quase em simultâneo em França e nos Estados Unidos (6,86), tendo este vindo a ser designado por vírus da imunodeficiência humana (HIV, do inglês ***Human Immunodeficiency Virus***).

O HIV é membro da família *Retroviridae* (Sub-Família *Orthoretrovirinae*), sendo incluído no Género *Lentivirus* e no Sub-Género dos Lentivírus dos primatas (59). É representado por dois tipos de vírus distintos: o HIV-1, que representa a espécie tipo dentro do género *Lentivirus* (86, <http://talk.ictvonline.org/>), e o HIV-2 (19). Estes, embora semelhantes, e ambos capazes de conduzir os indivíduos infectados a um estado de imunodeficiência profunda (característico da fase de SIDA) diferem, entre outros, nalguns aspectos da sua organização genómica, nas relações filogenéticas com outros vírus do mesmo Género, e apresentam cerca de 55% de divergência quando as suas sequências nucleotídicas são comparadas entre si (59). Enquanto o HIV-1 é responsável por uma pandemia a nível global, o HIV-2 encontra-se praticamente restrito a alguns países da África Ocidental, apresentado, ainda, prevalências muito reduzidas em alguns países Europeus (por exemplo em Portugal), em países Asiáticos como é o caso da Índia, e na África Oriental (como por exemplo em Moçambique) (82).

1.1 A síndrome da imunodeficiência adquirida

Em termos gerais, a infecção causada pelo HIV é caracterizada por um enfraquecimento continuado do sistema imunitário, de que um dos principais marcadores é a depleção de linfócitos T-auxiliadores (linfócitos T CD4⁺). Nela são

reconhecidas 3 fases principais. A primeira corresponde a uma infecção viral aguda, caracterizada por um aumento considerável da carga viral e a qual é acompanhada de uma diminuição abrupta do número de linfócitos T auxiliares. Informações obtidas com base em modelos animais (sobretudo utilizando macacos Rhesus infectados com vírus da imunodeficiência dos símios) sugerem que esta depleção inicial de linfócitos T $CD4^+$ deva ocorrer, principalmente, ao nível do sistema linfóide associado à mucosa intestinal (91). Esta primeira etapa termina com um decréscimo da carga viral à medida que resposta imunitária, especialmente devida à acção de linfócitos T citotóxicos $CD8^+$, se vai estabelecendo. A segunda fase é, essencialmente, uma fase assintomática, de duração variável, durante a qual ocorre a supressão progressiva da resposta imune, justificada através do aumento gradual da carga viral e acompanhada por um decréscimo irreversível do número de linfócitos T $CD4^+$. Finalmente, a terceira fase é sintomática e designada de SIDA. Esta está associada à falência terminal do sistema imunitário, com o número de linfócitos T $CD4^+$ a decrescer para níveis incapazes de sustentar a imunidade (13, 33) e culminando, geralmente, com o aparecimento de doenças provocadas pela multiplicação descontrolada de agentes patogénicos oportunistas, neoplasias e disfunção do sistema nervoso central (84, 101). As 3 fases da infecção por HIV encontram-se representadas na Figura 1.1.

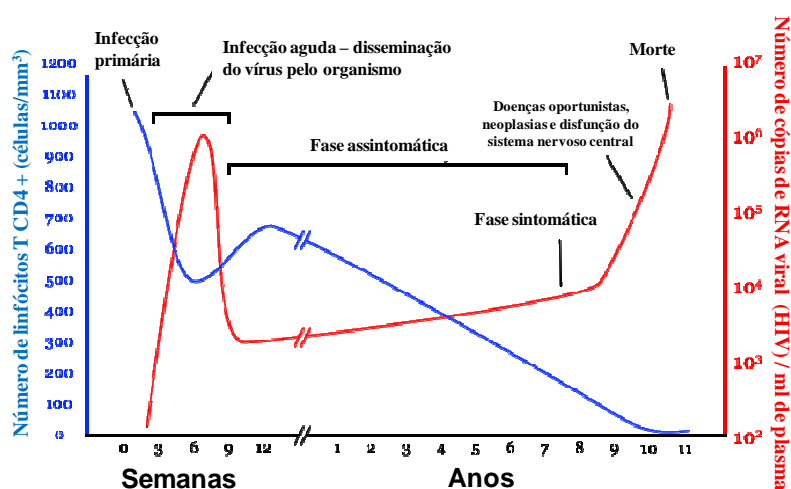


Figura 1.1 – Representação gráfica da história natural da infecção pelo HIV. Estão representadas as 3 fases principais: infecção aguda⁺, fase assintomática (latência clínica) e fase sintomática. São ainda apresentados os números de linfócitos (por μ l) e de cópias de RNA viral (por ml de plasma) característicos em cada uma das fases (adaptado de 77).

Rapidamente após a descrição dos primeiros casos de infecção pelo HIV, a sua dispersão assumiu um carácter epidémico. O último relatório da UNAIDS/WHO, datado de 2008, estima que, a nível mundial, cerca de 25 milhões de pessoas terão já falecido vítimas da infecção por HIV, devendo ainda existir cerca de 33 milhões de indivíduos infectados com este vírus, o qual será ainda responsável por cerca de 5 milhões de novas infecções anuais. Apesar do facto do HIV-1 se encontrar distribuído um pouco por todo o mundo, algumas regiões apresentam prevalências de infecção particularmente elevadas (Fig. 1.2). A África Sub-Sahariana continua a ser a região do Globo onde o problema da SIDA toma dimensões desproporcionadas, onde ocorrem 35% das infecções pelo HIV e 38% das mortes por SIDA, sendo também nesta região que habitam 67% dos indivíduos infectados com este vírus (113).

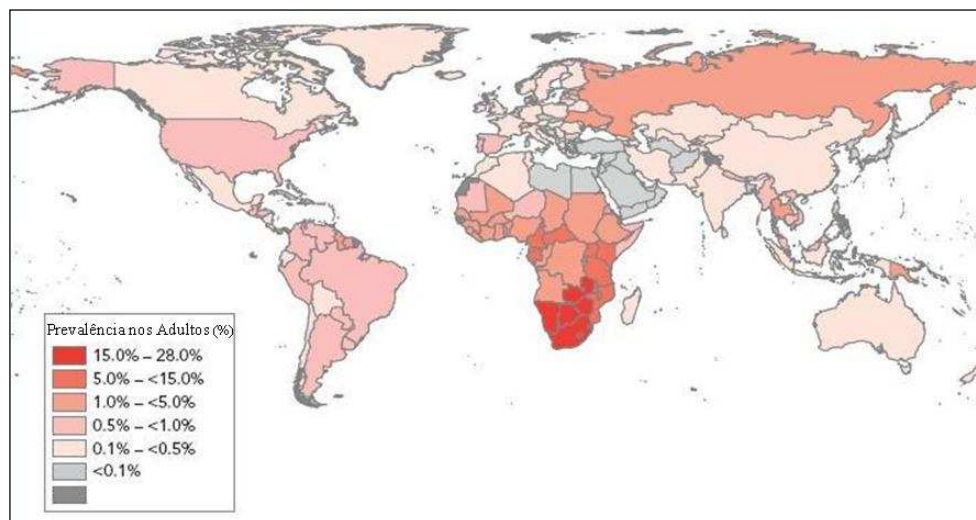


Figura 1.2 – Mapa representativo da variação na prevalência da infecção pelo HIV (do tipo 1 e 2) a nível Mundial (adaptado de 113).

1.2 O vírus da imunodeficiência humana do tipo 1

1.2.1 Estrutura morfológica do virião

Os viriões do HIV-1 correspondem a partículas com uma morfologia aproximadamente esférica e diâmetro a variar entre os 100 e os 120nm, revestidos por um invólucro lipídico que, normalmente, deriva da membrana citoplasmática das células hospedeiras onde replicam (Fig. 1.3). Esta camada exterior (também designada de invólucro) apresenta, aproximadamente, 72 trímeros de duas glicoproteínas, ambas codificadas pelo gene viral *env*. Estas estão associadas entre si de forma não-covalente e incluem um componente de superfície (SU/gp120) e um componente transmembranar (TM/gp41). O invólucro viral contém ainda proteínas celulares tais como os antígenos do sistema de histocompatibilidade (HLA) de classe I e II (59, 98). As glicoproteínas SU/gp120 desempenham funções importantes na interacção entre o virião e os receptores celulares, localizados à superfície da mesma. Por seu lado, as glicoproteínas TM/gp41 medeiam a fusão do invólucro do virião com a membrana da célula hospedeira. A face interna do invólucro viral é revestida por uma camada constituída pela proteína da matriz (MA, com 17kDa) que para além de desempenhar funções estruturais é importante no transporte de pré-integração para o núcleo (73). A cápside da partícula viral é constituída por uma estrutura proteica em forma de cone truncado, a qual é composta pela associação de monómeros da proteína da cápside (CA, de 24kDa). No interior da cápside localizam-se duas moléculas de ácido ribonucleico (RNA) viral, em associação com a proteína da nucleocápside (NC, de 7kDa). A cápside contém ainda as enzimas virais que incluem a transcriptase reversa (RT, heterodímero de duas proteínas de 51 e 66kDa), a integrase (IN, com 32kDa) e a protease (PR, com 10kDa), bem como as proteínas acessórias Vpr, Nef, Vif, e p6 (59, 98). A Figura 1.3 representa esquematicamente a estrutura morfológica dos viriões maduros de HIV-1.

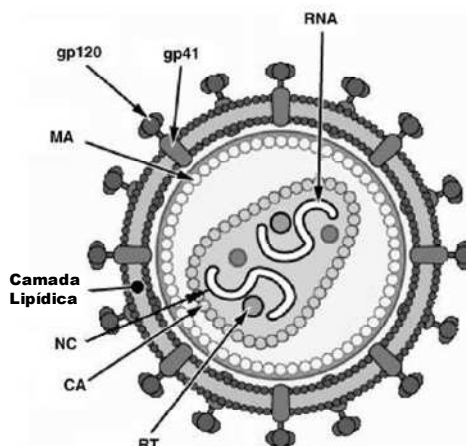


Figura 1.3 – Esquema representativo da estrutura do virião maduro de HIV-1. gp120 – glicoproteína de superfície; gp41 – glicoproteína transmembranar; MA – proteína da matriz (p17); CA – proteína da cápside (p24); NC – proteína da nucleocápside (p7); RT - transcriptase reversa (p51/66); RNA - ácido ribonucleico (adaptado de 98).

1.2.2 Organização do genoma viral e funções das proteínas virais

O genoma do HIV-1 é constituído por duas moléculas de RNA (genoma dimérico) de cadeia simples e polaridade positiva, com cerca de 10kb. Na extremidade 5' cada molécula de RNA genómico apresenta uma estrutura Cap metilada, enquanto que a extremidade 3' termina com uma cauda poliadenilada, características estas partilhadas com os mRNAs das células hospedeiras (20, 32). À semelhança dos restantes retrovírus, o genoma do HIV-1 é composto por três genes estruturais: *gag* (do inglês, *group-specific antigen*), *pol* (do inglês, *viral polymerase*), e *env* (do inglês, *envelope glycoprotein*). Para além destes, o genoma do HIV-1 contém ainda genes que codificam duas proteínas reguladoras (Tat e Rev) e quatro proteínas acessórias (Nef, Vif, Vpr e Vpu) (34, 99). Nos extremos do genoma viral existem sequências transcritas, mas não traduzidas, as quais, quando o genoma assume a sua forma proviral, imediatamente após a transcrição reversa, compõem as repetições terminais longas (LTR, do inglês, *Long Terminal Repeats*). Estas desempenham funções essencialmente reguladoras (Fig. 1.4).

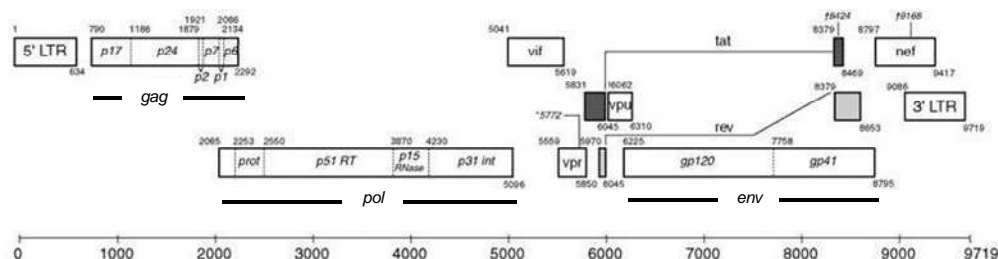


Figura 1.4 – Representação esquemática do genoma da estirpe HXB2 de HIV-1 (estirpe de referência). Estão identificados os genes constituintes, realçando os três genes estruturais (a sublinhado), bem como as proteínas que estes codificam. As coordenadas de cada uma das regiões são também indicadas (adaptado, <http://www.hiv.lanl.gov/>).

Os genes *gag* e *pol* codificam poliproteínas precursoras designadas Gag (de 55kDa) e Gag-Pol (ou Pr160, de 160kDa). Enquanto a primeira resulta da tradução directa da sequência do gene *gag*, a segunda é traduzida como resultando de um “salto” (*frameshift*) entre grelhas de leitura dos genes *gag* e *pol*, que ocorre durante a tradução de moléculas de RNA viral de tamanho completo. Ambas as proteínas são processadas, pós-tradução, pela protease viral, em proteínas distintas. O processamento exclusivo de Gag origina as proteínas MA, CA e NC, p6 e dos péptidos espaçadores p2 e p1, enquanto o processamento de Gag-Pol resulta, adicionalmente, na libertação das enzimas PR, RT e IN. A enzima RT assegura a transcrição reversa do genoma viral, convertendo o RNA viral em DNA proviral. Por seu lado, PR assegura a maturação dos viriões, enquanto, que IN é responsável pela integração do DNA proviral no genoma das células hospedeiras. O gene *env* codifica uma poliproteína precursora (gp160) sintetizada ao nível dos ribossomas associados ao retículo endoplasmático (ao contrário das demais proteínas virais). Esta é clivada, no aparelho de Golgi, por uma protease celular, durante o seu transporte para a membrana da célula, dando origem às glicoproteínas SU/gp120 e TM/gp41. O HIV-1 codifica duas proteínas reguladoras (Tat e Rev) que regulam a expressão dos genes virais a nível transcricional e pós-transcricional. A proteína Tat é responsável pela transactivação da transcrição iniciada na LTR do extremo 5', enquanto, que Rev desempenha, entre outros, um papel importante no transporte dos RNAs virais que ainda não sofreram *splicing* completo, do núcleo para o citoplasma (5, 41). Os restantes genes (*vpr*, *vpu*, *nef* e *vif*) codificam

proteínas acessórias, as quais desempenham funções importantes durante o ciclo replicativo viral. A proteína Vpu contribui para o processo de morfogénese e libertação da partícula viral da superfície celular, estando ainda implicada no mecanismo que conduz à degradação da proteína CD4 sintetizada *de novo*, contribuindo assim para a diminuição da sua abundância na superfície das células infectadas. A proteína Vpr desempenha várias funções que incluem a de transporte do complexo de pré-integração (ver secção 1.2.3) para o núcleo das células hospedeiras, a paragem do ciclo celular na fase G2, e a diminuição da transcrição a partir da LTR-5'. O gene *vif* codifica uma proteína que promove o aumento da infecciosidade viral, contribuindo para a depleção da proteína APOBEC3G da célula hospedeira (através da indução da sua degradação e da diminuição da respectiva taxa de tradução). Tal facto contribui não só de forma a prevenir um fenómeno designado de hipermutação G→A ou hiperadenilação, como também contraria o aumento da instabilidade das moléculas de RNA viral como resultado da criação de locais ocupados com nucleótidos sem base azotada, resultado da acção de Uracilo N-glicosilases incorporadas na partícula viral sobre os ribonucleótidos contendo uracilo que resultam da desaminação de nucleótidos de citidina pela APOBEC 3G. Finalmente, a proteína Nef é uma proteína à qual são atribuídas várias funções que incluem a estimulação da replicação viral, a supressão da apoptose nas células infectadas pelo HIV-1 e a indução de apoptose nas células vizinhas, a interferência na regulação da expressão das moléculas de CD4, CD8 e HLA de classe I à superfície das células infectadas e, ainda, a alteração de vias de activação celular (34, 35). As extremidades LTR, embora não codifiquem proteínas, são essenciais para a regulação da expressão génica viral, e incluem a região promotora, a região potenciadora (*enhancer*) e a região moduladora da transcrição (reconhecidas na LTR do extremo 5'), para além de incluir a sequência que, no extremo 3', é reconhecida pelas proteínas do hospedeiro responsáveis pela poliadenilação dos RNAs virais (35, 83).

1.2.3 Ciclo replicativo do HIV-1

O HIV é transmitido por via parentérica, essencialmente, através de contactos sexuais não protegidos (resultado da exposição das mucosas genital/anal), através da injeção de sangue (ou seus derivados) contaminado (incluindo a utilização de material de injeção não esterilizado), ou verticalmente, de mães infectadas para os seus filhos, durante a passagem pelo canal do parto ou em resultado de amamentação (<http://www.cdc.gov/>). Este vírus infecta células que apresentam receptores CD4⁺ na superfície, tais como linfócitos T auxiliares, monócitos, macrófagos, células de Langherans e da micróglia e células dendríticas. Este tropismo, para além de ser determinado pela utilização do receptor CD4⁺, é também devido à presença na superfície celular de moléculas co-receptoras, englobadas na grande família dos receptores de quimiocinas (essencialmente CCR5 e CXCR4) (18, 103).

O ciclo replicativo do HIV-1 apresenta-se esquematicamente representado na Figura 1.5, e inicia-se com a adsorção dos viriões a células susceptíveis. Esta decorre em etapas consecutivas que, segundo o modelo mais recente, envolve a endocitose da partícula viral pela interacção entre a glicoproteína gp120 com o receptor CD4⁺ e co-receptores. Assim, promoção da fusão do invólucro viral com a membrana celular, a qual depende de alterações conformacionais impostas a gp41, ocorre a nível intracelular (68, 112). O modelo aceite pressupõe, numa primeira fase, a ligação de gp120 ao receptor CD4⁺. Da interacção entre estas proteínas resultam alterações conformacionais em gp120 que conduzem à exposição do domínio de ligação aos co-receptores. Após a ligação de gp120 a estes últimos, novas alterações no trimero gp120/gp41 resultam na exposição do péptido de fusão localizado no extremo N-terminal de gp41 (83).

Uma vez no interior da célula hospedeira a partícula viral sofre descapsidação parcial e, em seguida, o RNA genómico é retro-transcrito pela RT. Este genoma viral é, então, transportado para o núcleo sob a forma de um complexo de pré-integração, no qual o genoma proviral (sob a forma de uma molécula de DNA de cadeia dupla) se encontra associado às proteínas virais MA, NC, IN, RT e Vpr, para além de diferentes proteínas celulares, seguido da sua integração no genoma da célula hospedeira (83).

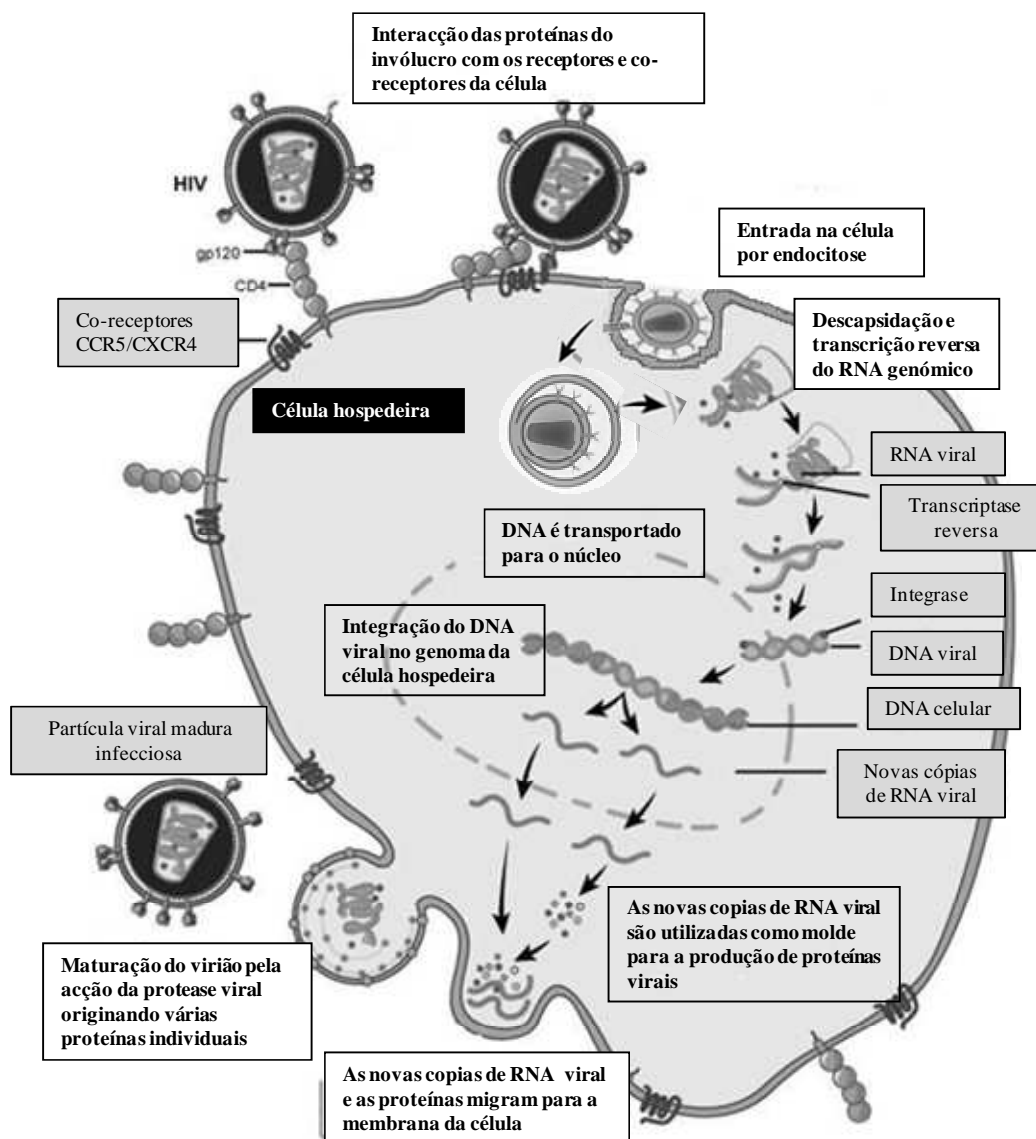


Figura 1.5 – Representação esquemática simplificada dos principais passos do ciclo replicativo do HIV-1 (adaptado de *National Institute of Allergy and Infectious Diseases (NIAID)*, disponível em <http://www.niaid.nih.gov>, 112).

Uma vez integrado, o DNA proviral serve de molde à síntese de mRNAs virais. A transcrição do genoma viral é levada a cabo pela RNA polimerase II celular e depende da ligação de factores de transcrição celulares a sequências reguladoras localizadas na LTR-5'. Os primeiros ciclos de transcrição (e consequente tradução) conduzem à síntese de quantidades basais das proteínas Tat, Rev e Nef, consequência do *splicing* múltiplo dos RNAs virais. Os RNAs resultantes são transportados para o

citoplasma, à semelhança do que acontece com os demais mRNAs celulares. À medida que os mRNAs se vão acumulando como resultado da transactivação da transcrição mediada por Tat, o transporte, para o citoplasma, dos mRNAs que não são sofrem *splicing* completo é assegurado pela actividade da proteína Rev. Estes mRNAs incluem os que servem de molde à tradução, pelos ribossomas citoplasmáticos, de Vif, Vpr e Vpu, bem como das glicoproteínas do invólucro ao nível do retículo endoplasmático (35, 98). O transporte de gp160 para a superfície celular é efectuado através da via secretora, durante o qual esta é glicosilada e clivada por uma protease celular (35). Trímeros de gp120/gp41 são levados, por vesículas secretórias, para a membrana citoplasmática da célula infectada onde, na maioria dos casos, ocorre a montagem dos viriões, e na qual participa a proteína p6. Domínios de resíduos de aminoácidos na proteína Gag são responsáveis pela encapsidação selectiva de RNA viral, bem como de um dos tRNA celulares (tRNA^{Lys3}), o qual serve de iniciador (*primer*) durante a reacção de transcrição reversa. A libertação das partículas virais, ainda na forma imatura não infecciosa, ocorre, normalmente, por gemulação e depende da inibição, mediada por Vpu, da actividade de uma proteína celular (BST-2, ou *tetherin*) que, de outra forma, mantêm os viriões presos à superfície celular (66). A partícula é, posteriormente, maturada como resultado da clivagem dos precursores proteicos (Gag e Gag-Pol) pela PR adquirindo, concomitantemente, infecciosidade (98).

1.3 Origem do HIV-1

Os estudos filogenéticos têm fornecido fortes evidências de que o HIV-1 teve a sua origem a partir de vírus da imunodeficiência dos símios (SIV) que infectam primatas não humanos Africanos (62). Apenas na década de 90 do século XX foi possível identificar a origem biológica do HIV-1, através da caracterização de uma sequência de um SIV dos chimpanzés (cpz) da subespécie *Pan troglodytes troglodytes* (SIVcpzptt), o qual revelou elevadas similaridades a nível genético com o HIV-1 (55, 37, 67). Tal sugestão é adicionalmente suportada pela elevada prevalência das infecções naturais por SIVcpzptt e sobreposição da área geográfica que engloba chimpanzés infectados por este vírus e humanos infectados por todos os grupos de HIV-1 (42).

Ao longo da literatura são descritos vários factores que contribuíram para a transmissão e evolução de carácter pandémico do HIV no princípio do século XX. A transmissão aos humanos poderá ter resultado da exposição da pele não intacta ou das mucosas humanas directamente ao sangue, ou secreções, de animais infectados, como resultado do consumo de carne de animais selvagens ou da sua utilização como animais de estimação (42, 67, 121, 122). Para além destes, outros factores como a urbanização, prostituição, ruptura social e o uso de material injectável não esterilizado deverão ter contribuído para a propagação rápida destes vírus nas populações africanas (42). Estes últimos factores foram evidenciados por alguns estudos que demonstram que o alastramento da epidemia coincidiu com o aparecimento de vastos núcleos populacionais nos países da África Central, facto este que se registou apenas depois de 1900 (96, 122).

1.4 Diversidade genética

1.4.1 Mecanismos de variação genética

O HIV-1 caracteriza-se por uma extraordinária diversidade genética. Entre os vários factores que para esta contribuem destaca-se a fraca fidelidade na síntese de DNA por parte da RT, uma vez que não apresenta actividade exonucleásica 3'→5'. Como consequência, em cada ciclo de replicação esta enzima permite, em média, a acumulação de cerca de $5,4 \times 10^{-5}$ mutações por nucleótido, o equivalente a 1,1 mutações em cada genoma por ciclo de infecção (36). Por outro lado, a presença de duas cópias do genoma na partícula viral possibilita a ocorrência de recombinação entre elas. Estima-se a existência de dois a trinta eventos de recombinação por genoma viral, em cada ciclo de replicação, dependendo estes, entre outros, do contexto celular (2, 49, 63). Em combinação com estes factores, a elevada taxa de produção de diária de viriões (cerca de 10^8 partículas virais) nos indivíduos infectados (4) contribui para o facto das populações virais serem geneticamente distintas, com diversidades genéticas na ordem dos 1% a 5%, sendo constituídas, no entanto, por estirpes virais geneticamente relacionadas. Estas, no seu conjunto, são designadas de *quasispécies* (4, 116).

1.4.2 Classificação das estirpes de HIV-1

Até recentemente, as estirpes conhecidas de HIV-1 apresentavam-se compreendidas em 3 grupos de vírus distintos: o grupo M (do inglês *Major*), O (do inglês *Outlier*) e o grupo N (não M e não O). Porém, recentemente foi identificada uma sequência viral classificada como pertencente a um novo, e potencial, grupo P, a qual exibe semelhanças com um vírus símio (SIVgor) encontrado em gorilas (*Gorilla gorilla gorilla*) (85) (Figura 1.6a). A título de exemplo, a análise das sequências de vírus dos grupos M e O revelaram uma divergência 30-50% para o gene *env* (102). Cada um destes 4 grupos de HIV-1 deverá ter resultado de um fenómeno de transmissão independente a partir de populações de primatas não humanos (37, 45, 55). Os vírus dos grupos O e N são responsáveis por menos de 5% do total de infecções. Os vírus do grupo O foram detectados, essencialmente, nos Camarões e países vizinhos da África Ocidental e Central, enquanto os vírus do grupo N foram identificados, exclusivamente, nos Camarões. Por seu lado, o grupo M inclui a grande maioria das estirpes de HIV-1 até hoje estudadas, sendo responsável pela pandemia a nível Mundial (62).

As sequências pertencentes a estirpes do grupo M agrupam-se em 9 subtipos (A-D, F-H, J e K, Figura 1.6b). Os subtipos estão relativamente equidistantes uns dos outros do ponto de vista filogenético, apresentando divergências entre si de cerca de 15% para o gene *gag* (8, 106), 13% para o gene *pol* e cerca de 25-35% para o gene *env* (9, 15, 81). São ainda reconhecidas categorias resultantes da subdivisão dos subtipos A (A1-A4) (117) e F (F1-F2), as quais são designadas de sub-subtipos (62).

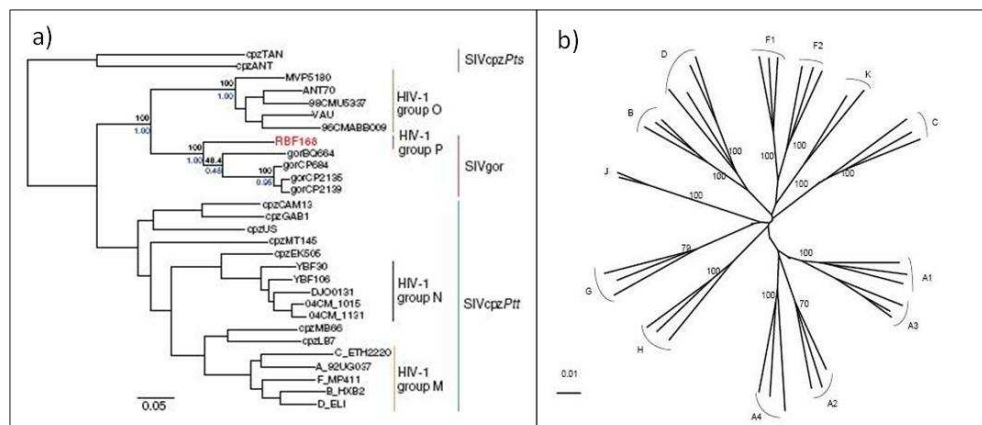


Figura 1.6 – Diversidade genética do HIV-1. a) Relações filogenéticas entre os lentivírus dos primatas, incluindo o novo grupo P de HIV-1 (adaptado de 85); b) diversidade genética do grupo M do HIV-1 (adaptado de www.retrovirology.com).

Como resultado da ocorrência de recombinação, muitas das estirpes virais identificadas apresentam um genoma que comporta regiões com identidade genética distinta. Estes vírus recombinantes, ou com genoma de tipo mosaico, são designados de formas recombinantes circulantes (CRF, do inglês *Circulating Recombinant Forms*) e sempre que sejam identificados num mínimo de três indivíduos epidemiologicamente não relacionados (106). Até à data, foram já descritas pelo menos 43 CRFs (<http://www.hiv-web.lanl.gov>, Janeiro de 2010). Os fenómenos de recombinação desempenham um papel importante na epidemiologia mundial do HIV-1, estando directamente relacionados com a diversidade e respectiva distribuição geográfica dos subtipos virais. As regiões de elevada diversidade genética (como é o caso da África sub-Sahariana) favorecem o aparecimento de formas recombinantes (9). Nestas regiões as formas recombinantes estão largamente distribuídas atingindo, em certos países, cerca de 60% do total de infecções, como é o caso da CRF02_AG nos Camarões e países vizinhos (9, 80). Pelo contrário, em regiões onde a epidemia é dominada quase exclusivamente por um único subtipo, existem uma menor probabilidade de surgirem vírus recombinantes (9). O fenómeno de globalização transformou a distribuição geográfica do HIV-1 num processo dinâmico e veio, de certa forma, aumentar a introdução de novos subtipos em regiões onde habitualmente não existiam, justificando, assim, o aparecimento de novos vírus recombinantes (80). A actividade

recombinogénica da RT viral não se esgota, no entanto, a identificação de CRFs, demonstra que a maioria dos vírus recombinantes são encontrados, em indivíduos epidemiologicamente relacionados (ver secção seguinte).

1.4.3 Distribuição geográfica global das estirpes HIV-1 do grupo M

Em termos globais, as variantes virais mais prevalentes são as classificadas nos subtipos C (47%) e A juntamente com CRF02_AG (27,2%), seguindo-se os vírus dos subtipos B (13,3%), D (5,3%) e CRF01_AE (3,2%) (74). A maior diversidade genética do HIV-1 é observada, no entanto, na região central Sub-Sahariana onde todos os subtipos são reportados. Ainda assim, os subtipos A e C são, em termos numéricos, os mais comuns (9, Figura 1.7). Estes últimos predominam ainda no Sul e Este africano, em países como a África do Sul, Etiópia, Tanzânia (106). Em particular na África do Sul vivem cerca de 5,7 milhões de indivíduos infectados pelo HIV, sendo o país apresenta a maior prevalência a nível Mundial (113). Devido à co-circulação local de variantes geneticamente distintas, o padrão epidemiológico relativo à infecção pelo HIV-1 é ainda caracterizado pela circulação de uma panóplia de formas recombinantes ditas únicas (URF, do inglês *Unique Recombinant Forms*), correspondentes a vírus recombinantes encontradas em indivíduos epidemiologicamente relacionados. São exemplo as URF_AC e URF_AD descritas em países como a Tanzânia, Zâmbia, Uganda e Quênia, onde circulam os subtipos A, C e D (51).

Ainda que o padrão epidemiológico relativo à distribuição dos subtipos de HIV-1 pelo Globo esteja em constante mudança, na Europa, América do Norte e Austrália, a maioria dos vírus circulantes pertencem, ainda hoje, ao subtipo B (Figura 1.7). Na América do Sul o subtipo B é também o mais frequente, contudo, os subtipos F e C, a CRF12_BF e URFs resultantes da recombinação entre vírus dos subtipos B/F têm sido frequentemente reportados (106). No Continente Asiático, nomeadamente na Índia, o subtipo C é o mais abundante, enquanto nas regiões do Sudeste Asiático são encontradas elevadas prevalências da CRF01_AE. Na Tailândia o subtipo B predomina, existindo uma variante única deste subtipo infectando a comunidade de utilizadores de

drogas injectáveis (UDI, Utilizadores de Drogas Injectáveis) (9). Na China têm sido identificados CRFs (CRF07_BC, CRF08_BC) epidemiologicamente associados à população de UDIs (57). Várias epidemias foram despoletadas pelas comunidades de UDIs, e da qual é exemplo a distribuição epidémica da CRF03_AB nas regiões orientais da Europa, em países como a Ucrânia, Bielorrússia e no exclave russo de Kaliningrado (69). Por outro lado, em algumas regiões da Europa a imigração tem contribuído para a introdução de estirpes genericamente designadas de não-B. A integração destes imigrantes na sociedade permitiu o aumento dos subtipos não-B nas populações nativas. A título de exemplo, na Inglaterra cerca de 23% das infecções por HIV-1 são atribuídas a vírus de subtipos não-B, infecções estas que são adquiridas no próprio país (3). Na Península Ibérica têm sido reportados vírus recombinantes entre o subtipo B e G, incluindo ainda o aparecimento da CRF14_BG (23). A semelhança de outros países da Europa que mantiveram laços históricos com África, Portugal apresenta elevadas taxas de infecções por vírus de subtipos não-B (cerca de 33%) na população nativa, em particular em utilizadores de drogas injectáveis (9).

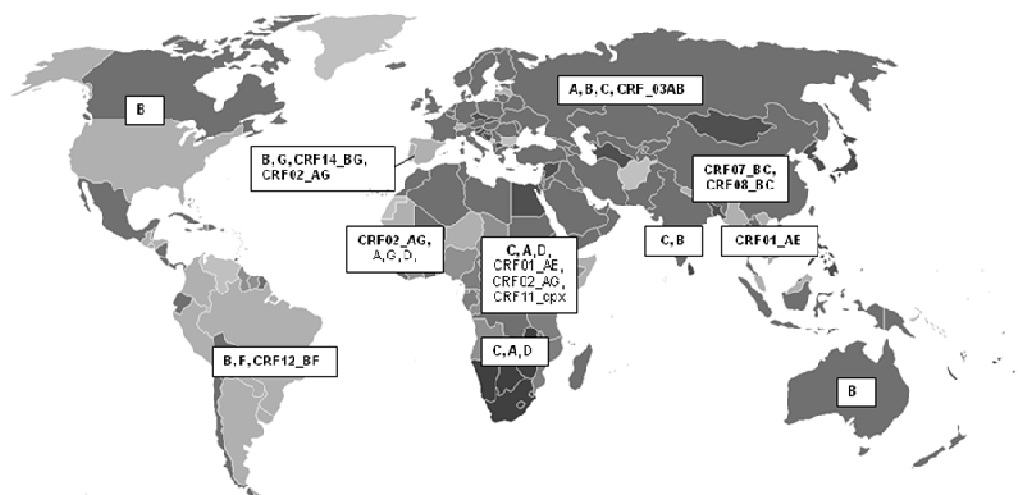


Figura 1.7 – Distribuição geográfica dos vários subtipos predominantes e formas recombinantes circulantes (CRFs) de HIV-1 do grupo M (adaptado de 106, 9).

1.4.4 Diversidade genética das estirpes de HIV-1 circulantes em Portugal

As estirpes de HIV-1 mais comuns em Portugal têm sido classificadas nos subtipos B (41,7 %) e G (29,4 %), contudo, ainda que menos frequentes, são reportados outros vírus classificados em inúmeros outros subtipos virais. São exemplos os vírus dos subtipos C (7,2 %), F (2,2 %), A (1,7 %), D (1,1) e J (0,6 %) (30, 75, 76, 78). Alguns estudos referem ainda prevalências na ordem dos 16% do total de infecções de formas recombinantes, nomeadamente CRF_02AG e CRF_14BG (Figura 1.7 e 1.8). São ainda distinguidas várias URF, maioritariamente resultantes de recombinação entre vírus dos subtipos B e G (76), frequentemente encontrados em grupos de risco como é o caso de dos consumidores de drogas injectáveis (29).

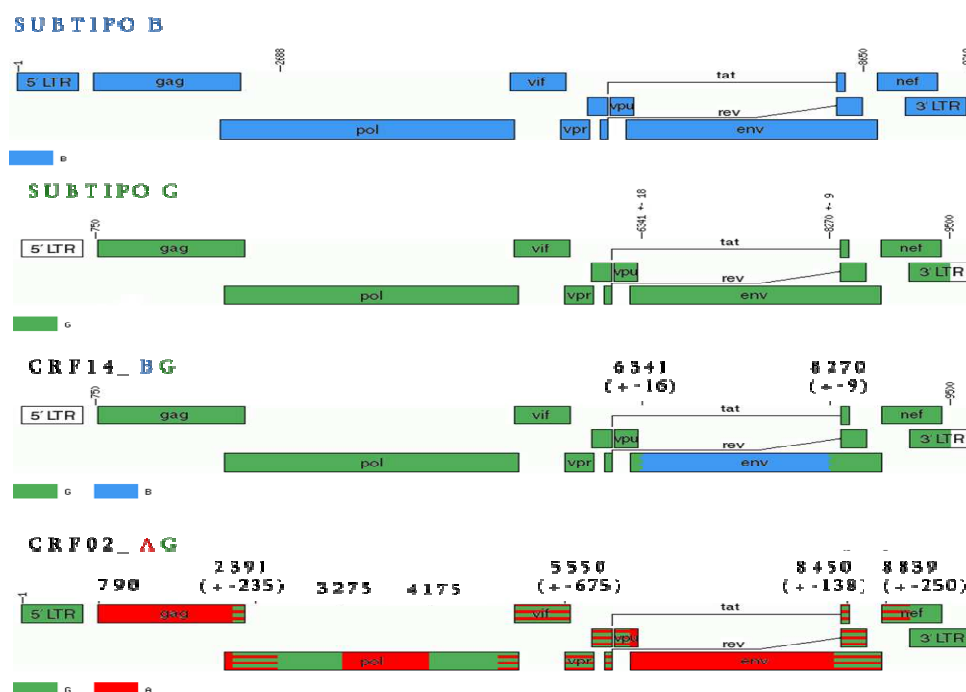


Figura 1.8 – Representação esquemática dos genomas das estirpes de HIV-1 mais frequentemente encontradas em Portugal. Estão também representados os pontos de recombinação das CRF02_AG e CRF14_BG, com as coordenadas referentes à estirpe de referência (HXB2) (adaptado de <http://www.hiv.lanl.gov/>).

A diversidade genética do HIV-1 em Portugal, evidenciada pela prevalência de vírus com secções do seu genoma (*gag* e/ou *env* e/ou *nef*) classificadas em subtipos não-B, está directamente relacionada com as ligações socio-económicas mantidas com as ex-colónias Portuguesas no continente africano (ex. Angola, Cabo Verde, Guiné Bissau ou Moçambique), na Ásia (Índia), e com a América do sul (Brasil) (76). Desta feita, existe um forte potencial para a contínua introdução de vários subtipos de HIV-1, decorrente do grande número de imigrantes que Portugal recebe anualmente (nos últimos anos incluindo, ainda, um grande número de indivíduos asiáticos e da Europa de leste). Para além destes, há ainda a considerar os portugueses que regressam ao nosso país após anos de trabalho, sobretudo, em África (30). O relatório de 31 de Dezembro de 2008 do núcleo de vigilância laboratorial de doenças infecciosas do Instituto Nacional de Saúde (INSA) refere a existência, até à data, de 34.888 casos de infecção por HIV/SIDA em Portugal. Em termos gerais, cerca de 42,5 % das infecções notificadas são referentes a UDI, 40 % envolvem casos de transmissão heterossexual, 12,3 % de transmissões por contactos homossexuais (masculinos) e 5,2 % implicam as restantes vias de transmissão. No entanto, os dados mais recentes demonstram uma alteração nos padrões de transmissão, com a transmissão por via heterossexual a evidenciar maiores prevalências (14). Este tipo de epidemia, associada a grupos de risco, tem um enorme impacto na diversidade genética do HIV, uma vez que potencia a ocorrência de co-infecções, permitindo eventos de recombinação e, consequentemente, aparecimento de novos vírus recombinantes.

1.5 Impacto da diversidade genética na epidemia do HIV-1

1.5.1 Patogenicidade, transmissibilidade e progressão para a doença

A variabilidade genética do HIV-1 é sem dúvida, um factor importante quando se avaliam as consequências da pandemia causada por este vírus. Os resultados de diversos estudos sugerem que os vários subtipos virais possam apresentar características biológicas distintas, tais como diferentes taxas de transmissibilidade, infecciosidade ou a progressão mais ou menos rápida para a doença (82, 106). Apesar disso, este assunto é

alvo de alguma polémica, resultante de dados contraditórios. A título de exemplo, um estudo na Tanzânia sugeriu que os vírus dos subtipos A, C e seus recombinantes apresentavam maior capacidade de transmissão perinatal do que os vírus do subtipo D, ao passo que no Quênia foi reportado que mulheres infectadas com o vírus do subtipo D apresentavam maiores taxas de transmissão vertical, comparativamente com as mulheres infectadas com o subtipo A (82). De uma forma geral, é difícil estabelecer uma relação inequívoca entre os subtipos e as diferenças nas transmissibilidade, patogenicidade e progressão para a doença, uma vez que estas podem não ser apenas dependentes de características dos vírus em questão, mas depender, também, de factores ambientais, como sejam a prevalência de outras infecções sexualmente transmissíveis e/ou parasíticas, do estado nutricional dos indivíduos infectados ou do seu património genético (106).

1.5.2 Resistência aos fármacos com actividade antirretroviral

Nos países desenvolvidos, onde predominam as infecções com os vírus do subtipo B, as terapias combinadas de alta eficácia, designadas por HAART (do inglês, *Highly Active AntiRetroviral Therapy*) têm originado uma enorme redução da transmissão, morbilidade e mortalidade associadas à infecção pelo HIV-1 (60,82). Estas abordagens terapêuticas fazem, em grande parte, uso combinado de dois grandes grupos de inibidores da RT: os análogos a nucleósidos (NRTI do inglês, *Nucleoside Reverse Transcriptase Inhibitors*) e os não análogos a nucleósidos (NNRTI do inglês *Non-Nucleoside Reverse Transcriptase Inhibitors*) (10). Para além destes, e essencialmente como drogas de segunda linha, podem ainda ser utilizados inibidores da protease viral e, em casos de falência terapêutica, inibidores da fusão, inibidores da entrada (inibidores de CCR5) ou inibidores da integrase (<http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/HIVandAIDSactivities/ucm118915.htm>). Em termos práticos, as terapêuticas combinadas permitem uma diminuição da carga viral, uma vez que limitam a replicação vírus e, consequentemente, restringem as probabilidades de transmissão (118), permitindo, ainda, o restabelecimento do sistema imunitário do indivíduo infectado (70). Apesar do sucesso terapêutico, e para além dos efeitos secundários associados, são frequentemente reportadas resistências aos

antirretrovirais utilizados. Estas resistências são suportadas por análises genotípicas, onde se identificaram diversas mutações nas sequências nucleotídicas referentes às regiões do genoma viral que codificam as proteínas alvo da acção antirretroviral (60). Estas mutações estão associadas, directa ou indirectamente, à selecção de variantes virais resistentes aos compostos especificamente associados a uma dada estratégia terapêutica ou, ainda, a toda uma classe de antirretrovirais (82, 102).

Alguns estudos realizados no âmbito da avaliação do impacto da introdução deste tipo de terapêutica em diferentes pontos do Globo onde predominam as infecções por vírus de subtipos não-B evidenciam diferentes graus de resposta aos antirretrovirais utilizados. Ainda que, *à priori*, as estirpes de todos os genótipos de HIV-1 sejam sensíveis aos compostos com actividade antirretroviral utilizados na terapêutica (o mesmo não acontecendo com alguns compostos apenas utilizados *in vitro*), os diferentes subtipos virais diferem na taxa com que são seleccionados vírus resistentes aos fármacos. De facto, análises genotípicas referem que, mesmo na ausência de terapia antirretroviral, algumas substituições nucleotídicas ocorrem naturalmente em determinados vírus, sobretudo nos classificados em subtipos não-B (polimorfismos naturais) em posições associadas a resistência antirretroviral nos vírus do subtipo B. Apesar disto, os poucos estudos efectuados em países africanos mostram que a taxa de falência terapêutica, associada à ocorrência de resistências aos antirretrovirais depende, maioritariamente, do seu uso inadequado (82). Em conjunto, a ocorrência de mutações/polimorfismos em determinadas posições do genoma viral, frequentemente associadas à implementação de terapia inadequada, ou de falha na adesão à mesma, estão na base da selecção das estirpes virais resistentes aos fármacos, reduzindo as estratégias terapêuticas disponíveis (70).

1.5.3 Desenvolvimento de vacinas

O desenvolvimento de uma vacina para o HIV-1 pressupõe, a longo prazo, o controlo da infecção ou da progressão da doença (SIDA). Considerando a diversidade genética do HIV-1, esta vacina deverá ser capaz de induzir imunidade protectora de forma a evitar as infecções por qualquer um dos subtipos, ou formas recombinantes, do

HIV-1 (82). Considerada a enorme variedade genética que caracteriza este vírus, esta tarefa toma, à partida, dimensões consideráveis.

A construção de uma vacina para o HIV-1 tem sido abordada de uma forma muito diversificada. De entre as estratégias utilizadas destacam-se as que fazem uso de vectores virais com base em Poxvírus (*canarypox*, *fowlpox* (varíola aviária) e estirpe *Ankara* do vírus da *vacina*), a administração de péptidos, proteínas, lipoproteínas ou se baseiam na utilização de vectores com origem no vírus da encefalite equina Venezuelana (88). De uma forma geral estas abordagens têm sido, maioritariamente, focadas nos vírus do subtipo B, resultado do facto de este ser o subtipo viral que regista prevalências mais elevadas nos países desenvolvidos (9). Contudo, em países onde os vírus não-B dominam a epidemia, e devido ao padrão não-uniforme de distribuição espacial dos diferentes subtipos virais e formas recombinantes, os esforços relativos ao desenvolvimento de uma nova vacina têm sido específicos para cada subtipo, dependendo da região (82). Algumas destas estratégias conduziram a testes clínicos em fase III. No entanto, a falta de eficácia e questões associadas à sua segurança, entre outros factores, têm impossibilitado, até ao momento, o desenvolvimento adequado de uma vacina para o HIV (88).

Em suma, uma vacina preventiva deverá induzir uma resposta humoral e celular, contudo, de modo a induzir respostas significativas para os vários subtipos, procura-se induzir respostas significativas, focalizando a atenção, particularmente, nas proteínas mais conservadas do HIV-1 (*gag* e *pol*). Este estudos demonstram que a reactividade cruzada observada quando a resposta imune é avaliada tendo em conta vírus dos vários subtipos varia consoante o grau de conservação do genes, sendo esta reactividade mais evidente dentro de um mesmo subtipo do que entre subtipos diferentes. Para além do mais, os resultados reportados na literatura são pouco claros ou mesmo contraditórios (62). Alguns estudos não estabelecem qualquer correlação entre os subtipos do HIV-1 e a resposta imunitária neutralizante, enquanto, que outros, realizados por exemplo na Ásia, sugerem esta correlação para os subtipos B e CRF01_AE (82).

Embora a grande maioria das estratégias de desenvolvimento de vacinas tenham sido implementadas a nível regional, a análise (extensa) das sequências de um grande número de estirpes do HIV-1, associadas à modelação computacional, poderá permitir a identificação de regiões nas diferentes proteínas virais com estrutura semelhante, ainda

que a partir de uma sequência primária diferente, resultando em novas estratégias de âmbito global (82), baseadas na utilização de uma sequência ancestral comum ou de um consenso de várias delas, o que poderá originar reacções cruzadas entre os vários subtipos e formas genéticas derivadas. Contudo, em regiões onde circulam vários subtipos a inclusão dos subtipos mais prevalentes numa vacina poderão aumentar a sua cobertura imunológica (9, 82).

1.5.4 Impacto da diversidade genética do HIV na sensibilidade dos métodos de diagnóstico e de monitorização da infecção (determinação de carga viral)

A diversidade genética do HIV tem um enorme impacto na sensibilidade dos métodos utilizados para o diagnóstico das infecções por ele causadas. No presente, o diagnóstico das infecções pelo HIV baseiam, na maioria dos casos, na utilização de testes serológicos, através dos quais se detectam anticorpos anti-HIV ou, no caso dos testes de 4ª geração, anticorpos anti-HIV e antígeno p24 (8, 62). Na sua grande maioria todos os testes comerciais permitem a detecção da maioria das estirpes virais pertencentes aos grupos M e O. Quer no seu formato imunoenzimático “clássico” (tipo ELISA; do inglês *Enzyme Linked Immuno Sorbent Assay*), quer no seu formato de “teste-rápido”, todos eles são sensíveis e específicos no diagnóstico de infecções estabelecidas. Contudo, para algumas infecções com subtipos não-B, resultados podem ser inconclusivos, justificados pelo facto de que os antígenos usados para os ensaios são baseados, sobretudo, na informação acumulada relativa ao subtipo B (62). Nos casos em que o diagnóstico serológico não pode ser aplicado (ex. diagnóstico precoce das infecções durante o período de “janela”, detecção de infecções por variantes genéticas pouco comuns, ou diagnóstico de infecções em crianças recém nascidas infectadas verticalmente), uma das estratégias possíveis para o diagnóstico das infecções por HIV baseia-se em abordagens moleculares (8, 62). Uma vez que estas pressupõem a detecção directa de sequências nucleotídicas das estirpes virais, frequentemente através da sua amplificação, *in vitro*, por reacção em cadeia da polimerase (PCR, do inglês *Polymerase Chain Reaction*) anteceder, ou não (dependendo do produto biológico de partida), de transcrição reversa (RT-PCR), o

impacto da diversidade genética viral torna-se óbvio (9, 62, 82). De facto, os conjuntos de *primers* genéricos utilizados têm de permitir a amplificação de todos os subtipos de HIV-1 e subsequente sequenciação dos mesmos, de modo a tornar possível a caracterização filogenética e identificação inequívoca dos diferentes subtipos virais.

Por seu lado, a detecção da carga viral é uma das abordagens experimentais que mais beneficiou do aperfeiçoamento dos ensaios moleculares, sendo particularmente importante para a avaliação terapêutica de indivíduos infectados pelo HIV. Actualmente, os métodos comerciais disponíveis dependem da detecção de sequências virais por oligonucleótidos através da amplificação das sequências alvo, ou do sinal emitido pelas sondas que se fixam ao RNA viral. Apesar das sequências alvo utilizadas corresponderem a regiões do genoma viral amplamente conservadas entre as diferentes estirpes virais, a elevada taxa de mutação associada aos fenómenos de recombinação têm tido impactos negativos na sensibilidade de detecção pretendida, especialmente quando utilizados ensaios comerciais de fabricantes distintos (62). Apesar dos muitos avanços nas técnicas moleculares de genotipagem e quantificação de carga viral, são necessários ajustes constantes nas mesmas, uma vez que a distribuição dos subtipos e emergência de novas estirpes (incluindo recombinantes) é um processo dinâmico, em especial nos países menos desenvolvidos do continente africano (62, 82).

1.6 Métodos de genotipagem de HIV-1

São vários os métodos que, ao longo do tempo, têm vindo a ser desenvolvidos de forma a permitirem análise da diversidade genética que caracteriza o HIV-1. Inicialmente, e com o objectivo de subtipar um elevado número de amostras com custos relativamente reduzidos, desenvolveram-se estratégias alternativas à sequenciação do genoma (7). Embora as estratégias de genotipagem sejam distintas entre si, a maioria dos métodos contemplam uma etapa inicial comum. Esta etapa caracteriza-se pela extracção, isolamento e transcrição reversa do RNA viral, a qual é assegurada por transcriptases reversas obtidas comercialmente. Como alternativa, esta genotipagem pode basear-se na amplificação de DNA proviral a partir de extractos de DNA de células infectadas.

1.6.1 Genotipagem de HIV-1 baseados na análise da mobilidade diferencial de homo e heteroduplexes

Um dos métodos amplamente utilizados para a genotipagem de HIV-1 foi designado de ensaio de mobilidade de heteroduplexes, vulgarmente conhecido como HMA (do inglês *Heteroduplex Mobility Assay*) (25). O ensaio de HMA baseia-se na análise da mobilidade diferencial das moléculas de DNA híbridas (heteroduplexes) que se formam entre um fragmento de DNA, resultante da amplificação de uma qualquer estirpe viral a subtipar, e a mesma região amplificada a partir de uma amostra de referência. Os heteroduplexes são obtidos através da mistura de dois produtos de PCR, da sua subsequente desnaturação a elevadas temperaturas (normalmente 94°C durante alguns minutos) seguida de um arrefecimento rápido permitindo, assim, a renaturação das moléculas de DNA. Para além dos heteroduplexes que se formam quando uma cadeia polinucleotídica da amostra a subtipar híbrida com a complementar da amostra de referência utilizada, este arrefecimento rápido promove ainda a formação de homoduplexes (quando as duas cadeias exactamente complementares hibridam entre si) e moléculas que permanecem em cadeia simples. O número de reacções individuais de hibridação depende do número de estirpes de referência disponíveis e/ou para as quais se quer testar a amostra a subtipar. A identificação das moléculas híbridas é realizada na presença de brometo de etídio (sob luz UV) após a separação electroforética em gel de poliacrilamida (5%) das moléculas presentes em todas as misturas de hibridação. Este tipo de ensaio recorre à análise de duas regiões dos genes *gag* e *env* do HIV-1 (24, 26, 114). O esquema a baixo (Figura 1.9) ilustra, de forma resumida, os principais passos realizados no decurso desta estratégia.

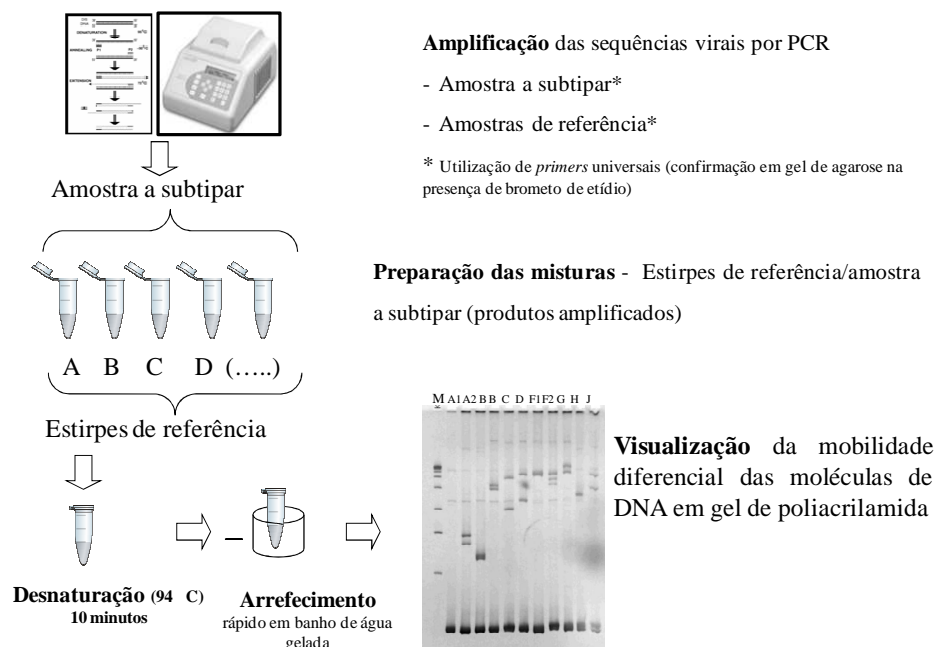


Figura 1.9 – Esquema representativo dos pontos-chave de um ensaio de HMA para subtipagem do HIV-1. Inclui: **Amplificação** por PCR (ou PCR em tempo-real) dos fragmentos de DNA da amostra a subtipar (regiões *gag* e/ou *env*) e das amostras de referência (PCR); Preparação das **misturas de reação** (amostra a subtipar/amostras de referência); **Desnaturação** a 94°C das misturas de reação; **Hibridação** por arrefecimento em num banho de água gelada; **Migração** em gel de poliacrilamida e identificação dos homo/heteroduplexes.

Uma vez que a identidade genética determina a conformação dos híbridos de DNA formados, a qual, por sua vez, condiciona directamente a sua migração electroforética em gel de poliacrilamida, quanto mais esta se aproximar da dos homoduplexes, maior a semelhança das duas cadeias polinucleotídicas hibridadas sob a forma de um heteroduplex. O subtipo viral é indicado, então, pelo subtipo da amostra de referência cuja hibridação com a amostra a subtipar resultou na identificação de heteroduplexes com a maior mobilidade electroforética (26).

Um grupo de investigadores desenvolveu um método que permite a subtipagem das regiões p24 e p7 do gene *gag* e C2-V5 do gene *env*, também baseado na análise de homo e heteroduplexes em gel de poliacrilamida, o qual foi denominado de rHMA (do inglês *reverse HMA*, (7). Apesar de tecnicamente muito semelhante ao método anterior,

ao invés de utilizar um conjunto de estirpes de referência abrangente, tanto quanto possível, da diversidade genética viral em termos globais, no rHMA são utilizadas apenas estirpes de referência do genótipo dominante (ou dos dois mais representativos) na área geográfica onde se pretende implementar o método. A análise das estirpes virais é efectuada através do cálculo de mobilidades relativas, correspondendo estas à distância migrada pelos heteroduplexes relativamente à distância percorrida pelos homoduplex em “mini”-gel, seguida da sua comparação com valores padrão (7). Este conjunto de alterações permite rastrear um maior número de amostras, uma vez que o número de combinações (amostra a subtipar/amostra de referência) e o tempo requerido para a execução do mesmo (condicionado pelas dimensões do gel de poliacrilamida) é inferior ao do HMA convencional. Na sequência da utilização deste tipo de abordagem, as estirpes virais (incluindo formas recombinantes) que apresentam resultados não discriminatórios seriam posteriormente analisadas (7). No entanto, este método poderá apresentar limitações de implementação em regiões onde a epidemia do HIV-1 apresenta várias estirpes circulantes, comparativamente à versão original (HMA).

Um outro método que partilha algumas semelhanças técnicas com o HMA é o denominado de HTA (do inglês *Heteroduplex Tracking Assay*). Neste caso as misturas de reacção são compostas pelos produtos amplificados *in vitro* (DNA da amostra a subtipar) e de uma sonda específica marcada radioactivamente (para cada subtipo ou forma recombinante) (72). A visualização dos heteroduplexes é efectuada por autorradiografia após migração e desidratação do gel de acrilamida. Este método foi desenvolvido para a região V3 do gene *env*, revelando-se como um método de alta resolução permitindo, para além da subtipagem *per se*, a identificação de várias populações de vírus de um mesmo subtipo num indivíduo infectado. A análise comparativa entre os padrões de mobilidade e as respectivas sequências demonstrou que apenas duas alterações na cadeia justificam uma alteração na mobilidade dos heteroduplexes (72). Embora seja uma estratégia altamente sensível, o tempo necessário para a migração do gel de acrilamida (cerca de 6 horas), as inúmeras combinações necessárias para cobrir todas as variantes genéticas do HIV-1, e ainda a manipulação de material radioactivo, criam algumas limitações à sua implementação.

1.6.2 Métodos de genotipagem de estirpes HIV-1 recorrendo à incorporação de moléculas fluorescentes nas cadeias de DNA

Para além dos métodos apresentados anteriormente existem ainda um conjunto de métodos de genotipagem de HIV-1 descritos na bibliografia (desenvolvido pela *Specialty Laboratories, Inc*, EUA, referido em 120), de que são exemplos o GenotypeR e GeneChip(RT) (desenvolvido pela Affymex, EUA, referido em 120). O GenotypeR, em termos gerais, compreende um protocolo de sequenciação (método de Sanger) modificado, uma vez que os amplicões (produtos de amplificação) produzidos incluem nucleótidos previamente marcados com moléculas fluorescentes (uma diferente para cada nucleótido). Estes amplicões são sujeitos a uma migração em gel de poliacrilamida. À medida que ocorre a migração, os nucleótidos são excitados individualmente por acção de um feixe de luz monocromática (laser), sendo o sinal luminoso traduzido por um *software* informático originando uma sequência nucleotídica virtual. Esta sequência é posteriormente comparada com uma sequência consenso permitindo a genotipagem da região em análise (referido em 120). No caso do GeneChip (RT) a sua utilização baseia-se na incorporação de inúmeras (mais de 16.000) matrizes complementares a regiões ao genoma viral (nesse caso regiões da RT) num chip. Estas matrizes são utilizadas como sondas que vão hibridar com fragmentos de ácidos nucleicos previamente marcados com moléculas fluorescentes. A fluorescência é detectada por um equipamento de leitura óptica e posteriormente traduzida pelo *software* informático. Embora estas duas últimas estratégias tenham sido desenvolvidas com a finalidade de genotipar estirpes clínicas de HIV, e não obstante o facto de que para a sua implementação seja necessário a utilização equipamentos e aplicações informáticas específicas, a sua utilização prática restringe-se à detecção de resistências aos antirretrovirais. Este facto determina que as regiões do genoma analisadas sejam apenas as que constituem alvos terapêuticos (120).

Embora todos os métodos até agora descritos permitam a genotipagem parcial de estirpes de HIV, no máximo inspeccionando duas regiões do genoma (*gag* e/ou *env*), a sua utilização está longe de permitir avaliar as características do genoma na sua totalidade (56). De uma forma geral, a experiência tem sistematicamente demonstrado, que, a aplicação, de estratégias que contemplam análises parciais, do genoma viral

originam muito mais frequentemente do que inicialmente previsto, resultados incorrectos, na genotipagem das estirpes de HIV circulantes. Esta situação agrava-se, principalmente, em regiões geográficas de maior heterogeneidade do HIV, onde podem co-existir várias formas virais circulantes e onde, por conseguinte, os eventos de recombinação e/ou a ocorrência de infecções mistas são potenciadas (47). Para além desta limitação, a inexistência, para alguns dos métodos, como é o caso do HMA, de estirpes de referência que permitam cobrir a maior parte da diversidade genética do HIV-1, levanta alguns condicionantes do ponto de vista analítico (47, 56).

1.6.3 Genotipagem de HIV-1 por sequenciação do genoma viral

Entre outros, os métodos de genotipagem referidos nas secções anteriores revelaram dificuldades em garantir uma correcta genotipagem das estirpes do HIV-1. Perante esta realidade, a sequenciação completa do genoma viral continua a ser o único método capaz de fornecer uma informação exhaustiva, e como um todo, da natureza genética das estirpes circulantes. Na bibliografia existe um conjunto de estratégias que permitem a sequenciação do genoma em toda a sua extensão. Uma das mais utilizadas baseia-se na realização de sucessivas reacções de PCR (*hemi* ou *nested-PCR*), das quais resultam vários amplicões sobreponíveis, e os quais permitem cobrir toda a extensão do genoma (Figura 1.10) (71, 109).

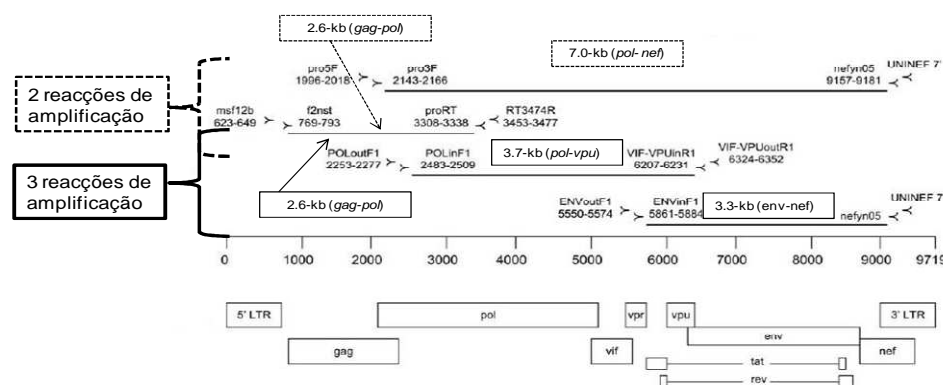


Figura 1.10 – Representação esquemática de um protocolo experimental para a amplificação do genoma completo de HIV-1. Estão representadas duas estratégias: a primeira contempla uma reacção de *nested-*

PCR, a qual origina dois fragmentos sobreponíveis com 2,6kb (*gag-pol*) e 7,0kb (*pol-nef*); a segunda contempla 2 reacções de *nested*-PCR repartindo o fragmento de 7,0kb da estratégia anterior em dois fragmentos individuais, com 3,7kb (*pol-vpu*) e outro com 3,3kb (*env-nef*) (adaptado de 71).

Uma vez obtidos, todos os amplicões são sequenciados individualmente e os resultados editados sob a forma de numa sequência contígua (frequentemente designada de *contig*). A genotipagem do genoma viral como um todo é levada a cabo por inferência filogenética (Figura 1.11a), através da análise comparativa das sequências obtidas com sequências nucleotídicas correspondentes a genomas completos de estirpes virais de referência (71, 109). Este método apresenta-se como um procedimento extremamente dispendioso, laborioso e de difícil implementação na rotina laboratorial, especialmente para um elevado número de amostras e na ausência de infra-estruturas que permitam reduzir a carga de trabalho envolvida recorrendo a métodos totalmente automatizados. Assim sendo, a estratégia mais utilizada, apresentada como referência, consiste na sequenciação e análise filogenética de várias regiões do genoma (51). Este tipo de abordagem, para além de permitir determinar o subtipo de cada uma das regiões individualmente consideradas (Figura 1.11b), dá ainda lugar à avaliação das relações filogenéticas entre as várias sequências virais analisadas para cada região amplificada (51, 107). Apesar de estratégia nos permitir identificar sequências recombinantes (quando duas regiões do genoma amplificadas da mesma estirpe viral são classificadas em subtipos distintos), a sua identificação pode ser confundida com a possibilidade de ocorrência de infecções mistas. Neste caso, a presença simultânea, numa dada amostra biológica, de diferentes estirpes virais com origens genéticas distintas, pode gerar resultados inconclusivos ou interpretados como correspondentes a vírus recombinantes.

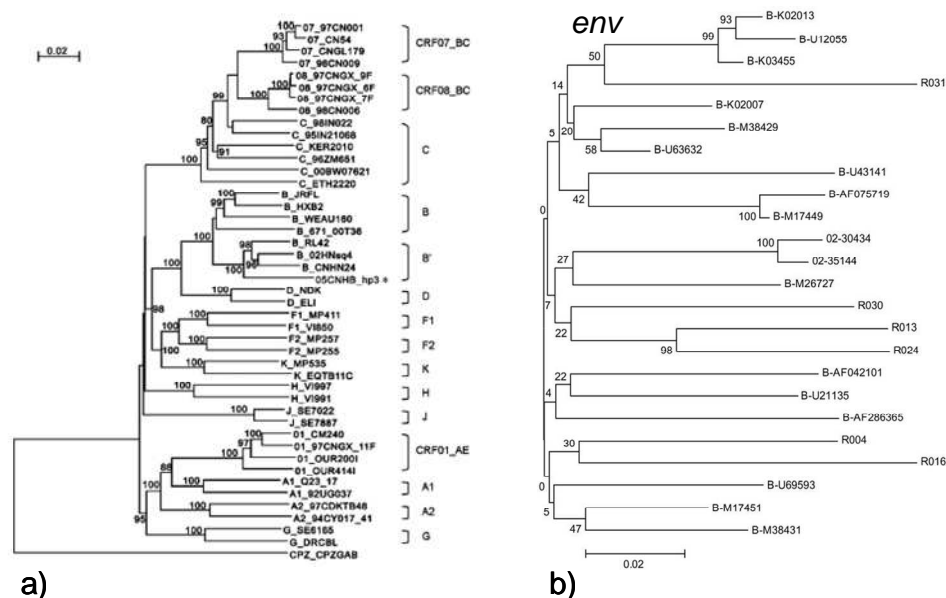


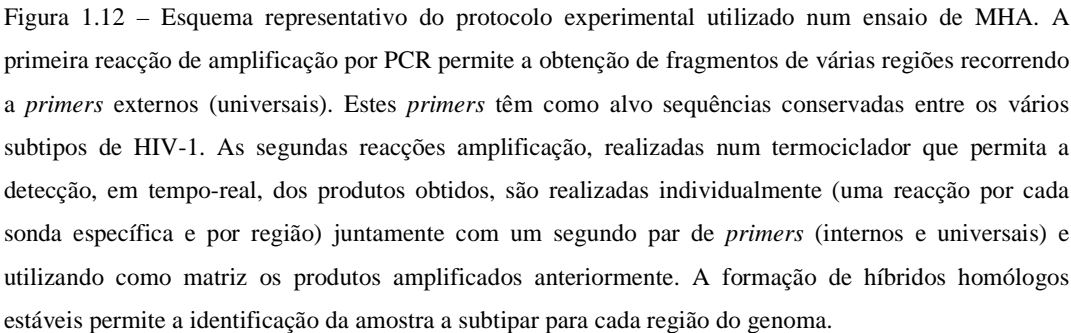
Figura 1.11 – Esquema de árvores filogenéticas para a análise filogenética do HIV-1: a) exemplo de uma árvore filogenética típica gerada a partir da análise de sequências de genoma completo (adaptado de 105); b) exemplo de uma árvore filogenética construída a partir da análise de uma região do genoma, neste caso a região C2V3 do gene *env* (adaptado de 105).

1.6.4 Genotipagem de HIV-1 através da utilização de métodos de hibridação múltipla

Com o objectivo de manter a especificidade e, simultaneamente, apresentar uma estratégia relativamente menos dispendiosa e mais rápida que a sequenciação do genoma, foi desenvolvido um novo grupo de métodos para a genotipagem simultânea de diferentes regiões do genoma de HIV-1, genericamente designado de hibridação múltipla (MHA do inglês, *Multiregion Hybridization Assays*) (47). Ainda que a identidade genética do genoma viral não seja avaliada como um todo, o objectivo destes ensaios é o de fornecer, para cada estirpe, informação genética suficiente, cobrindo várias regiões do genoma em simultâneo, de forma a distinguir os subtipos e formas recombinantes que circulem numa dada região geográfica. Estes ensaios são, normalmente, desenvolvidos a nível regional, uma vez que as estirpes circulantes numa

dada área podem, como mencionado anteriormente, diferir das reportadas noutras regiões geográficas (57).

O princípio geral do MHA baseia-se na amplificação individual, por PCR/ PCR em tempo-real de várias regiões ao longo do genoma do HIV-1, em duas reacções consecutivas, e na hibridação destes fragmentos amplificados com sondas fluorescentes específicas, numa plataforma de PCR em tempo-real (47). A primeira reacção de amplificação permite a obtenção de um fragmento de DNA recorrendo a *primers* sintéticos denominados de *primers* externos. Estes são desenhados de forma a permitirem a amplificação *in vitro* de sequências virais independentemente das características genéticas da estirpe viral em questão, e justificando a sua designação como “*primers* universais”. Por sua vez, os fragmentos de DNA obtidos na primeira reacção servem de matriz para uma segunda reacção amplificação (PCR em tempo-real) utilizando um segundo par de *primers* denominados de *primers* internos, os quais, à semelhança dos anteriores, são universais. A especificidade do ensaio é garantida pela utilização de uma sonda subtipo/CRF-específica geralmente do tipo *TaqMan* (56, 57) nas segundas reacções de PCR (Figura 1.12). Em teoria, os *primers* denominados de externos e internos permitem a amplificação de todos os subtipos e formas recombinantes do HIV-1 circulantes numa qualquer área geográfica. As regiões de hibridação das sondas específicas estão situadas internamente às regiões alvo dos *primers* internos. Idealmente, e quando são conhecidas estirpes recombinantes (CRFs) na epidemia local, os fragmentos amplificados devem estar situados em regiões do genoma que flanqueiem os pontos de recombinação dos CRFs em questão (57). A Figura 1.12 representa esquematicamente o protocolo experimental típico deste tipo de ensaios.



Como mencionado anteriormente, a grande maioria dos ensaios de MHA faz uso, na etapa de PCR em tempo-real, de sondas do tipo *TaqMan* (46, 47, 56, 57, 119). Estas moléculas são genericamente conhecidas como sondas de hidrólise uma vez que a fluorescência por elas emitida é consequência da sua degradação. De facto, a fluorescência, emitida por um fluoróforo covalentemente ligado ao extremo 5' da

mesma, é inicialmente suprimida pela presença, *in cis*, de um “abafador de sinal” (do inglês *quencher*), ligado covalentemente ao extremo 3’ da sonda. Aquando da reacção de amplificação, a actividade exonucleásica 5’→3’ da polimerase utilizada promove a hidrólise desta sonda (previamente hibridada ao DNA a amplificar), garante a separação espacial entre o fluoróforo e o *quencher*, e permite que o aparelho de PCR detecte o sinal de fluorescência emitido pelo primeiro (65) (Fig. 1.13). Estão disponíveis comercialmente várias moléculas (fluoróforos e “abafadores de sinal”) acopláveis às diferentes sondas nos seus diferentes formatos.

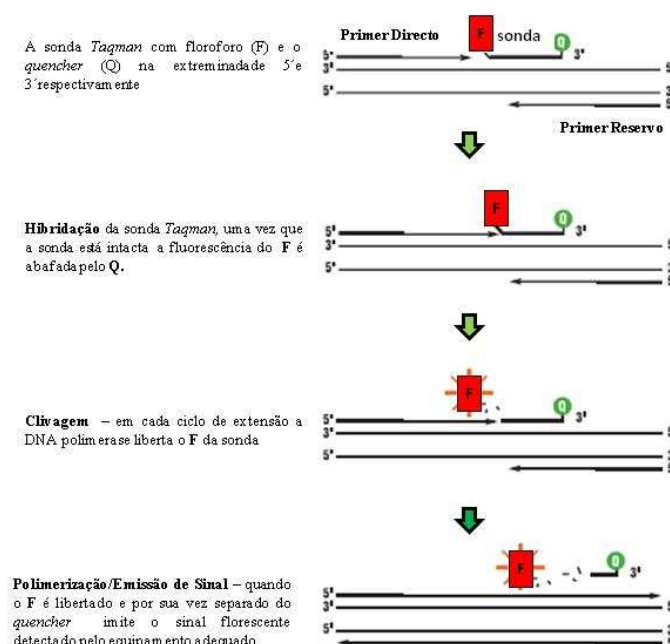


Figura 1.13 – Representação esquemática genérica do sistema envolvido na detecção de sinal fluorescente emitido como resultado da utilização de sondas tipo *TaqMan* numa plataforma de PCR em tempo-real (adaptado, www3.appliedbiosystems.com).

O sucesso da utilização deste método depende, especificamente, do delineamento dos *primers* universais e sondas específicas utilizadas. Por um lado, os primeiros deverão permitir amplificações que não dependam da identidade genética da estirpe viral em causa. Por outro, a sequência nucleotídica das sondas deverá permitir uma óptima hibridação com as sequências homólogas (dos respectivos subtipos/CRF) e, simultaneamente, reduzir ao máximo a possibilidade de hibridação cruzada com estirpes

heterólogos (57). De uma forma geral esta estratégia requer que se cumpram alguns critérios. Idealmente, os *primers* universais utilizados devem apresentar temperaturas de desnaturação compreendidas entre os 58-60°C, ao passo que esta temperatura deve ser superior para as sondas utilizadas (61-68°C). Por outro lado os *primers* devem compreender uma sequência nucleotídica variável entre os 15 a 30 nucleótidos e apresentar um conteúdo C/G a variar entre os 30 a 60%. Deve ser evitada a utilização de sondas que apresentem vários nucleótidos G/C nas extremidades ou repetições dos mesmos ao longo da sequência, uma vez que promovem a instabilidade e hibridações inespecíficas. Finalmente, e para que o processo seja o mais específico possível, os amplicões gerados não devem exceder os 400pb (<http://www.premierbiosoft.com>). Alguns estudos sugerem, ainda, que para a formação de híbridos estáveis entre a sonda e os seus alvos, em detrimento da formação de híbridos heterólogos às temperaturas de hibridação e polimerização utilizadas, é necessário que existam, pelo menos, 3 nucleótidos distintos (não homólogos) entre as várias sondas a utilizar. Contudo, com o objectivo de aumentar as diferenças poderão ser introduzidos nucleótidos não homólogos na sequência da sonda, os quais desfavorecem a formação de híbridos heterólogos (47, 56, 57) (Figura 1.14).

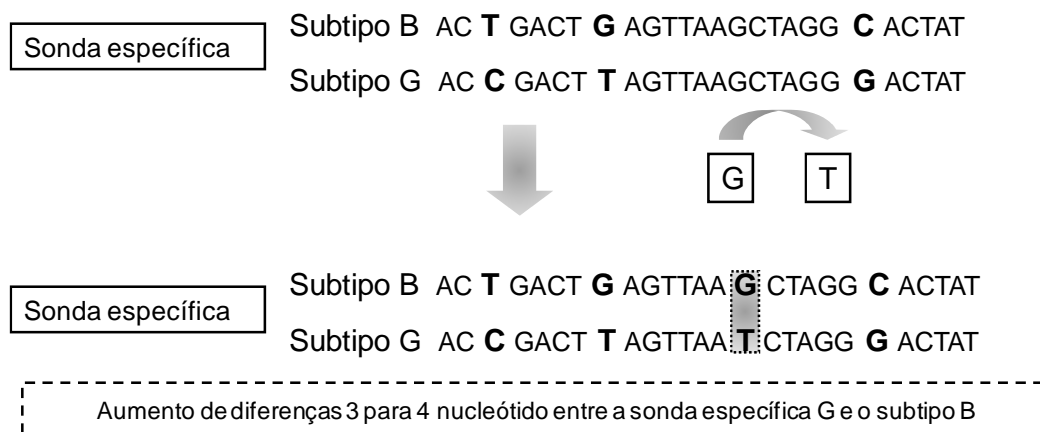


Figura 1.14 – Esquema representativo da construção de sondas específicas a implementar em ensaios de MHA. Neste esquema são representadas sondas para dois subtipos B e G, numa região alvo onde existem naturalmente 3 diferenças nucleotídicas (a negrito) quando as sequências de vírus dos dois subtipos são comparadas entre si. Foi introduzida propositalmente uma substituição de um nucleótido de modo a desfavorecer a ligação entre a sonda G e o subtipo B (assinalada a sombreado).

Embora seja possível prever a estabilidade térmica dos híbridos alvo/sonda com base na extensão e composição da sequência nucleotídica da sonda utilizando, para tal, algoritmos matemáticos, a estabilidade da hibridação encontra-se sempre dependente da natureza da sequência alvo e da concentração dos reagentes utilizados (111).

Em suma, com este tipo estratégia é possível, em teoria, genotipar várias regiões do genoma viral em simultâneo, permitindo a detecção de possíveis formas recombinantes. Comparativamente à sequenciação do genoma completo, uma vez otimizados, estes ensaios apresentam algumas vantagens, na medida em que são menos dispendiosos e de execução relativamente fácil, permitindo a genotipagem rápida e em larga escala das estirpes virais circulantes numa dada população. A literatura recente revela já o desenvolvimento e implementação deste tipo de métodos em várias regiões do Globo, distintas nas características genéticas das estirpes virais circulantes, evidenciando bons resultados. A título de exemplo o MHA desenvolvido para as regiões do Oeste e centro Ocidentais de África, onde esta metodologia foi utilizada para a identificação do CRF02_AG. Este estudo refere valores elevados de especificidade e sensibilidade (100%) a partir da utilização de DNA de células mononucleadas de sangue periférico (PBMC, do inglês *Peripheral Blood Mononuclear Cells*) (56). Ensaios semelhantes foram desenvolvidos para serem aplicados na África Oriental de forma a permitir a identificação indubitável de vírus dos subtipos A, C e D e suas formas recombinantes (47). De igual modo esta abordagem foi já desenvolvida e aplicada na Ásia com o intuito de identificar alguns dos subtipos mais abundantes nesta região, como sejam os subtipos B, C e a CRF01_AE (57), ou ainda na América do Sul (Perú) para identificação dos subtipos B, F e suas formas recombinantes (46).

1.7 Notas finais e objectivos

Para além das inúmeras questões que continuam por responder sobre inúmeros aspectos da biologia do HIV, a diversidade genética e as potenciais e particulares características associadas aos diferentes subtipos virais, onde se incluem os determinantes da sua evolução epidemiológica no espaço e no tempo, continuam a impor vários desafios no campo da genotipagem. A grande maioria dos métodos de caracterização genética do HIV-1 baseia-se em análises singulares e parciais de regiões específicas do genoma, revelando-se ineficazes principalmente em virtude da existência de recombinações entre os vários subtipos e CRFs. Se bem que a sequenciação do genoma viral completo seja o método que garante a melhor caracterização genética das estirpes virais circulantes, permitindo, ainda, a detecção de genomas recombinantes, ela apresenta-se como um método dispendioso, tecnicamente exigente e laborioso, sendo, por isto, difícil de implementar em larga escala num curto espaço de tempo. Tendo em conta as alternativas possíveis, este estudo foi levado a cabo com base em dois objectivos principais. Por um lado pretendeu-se desenvolver um novo ensaio de hibridação múltipla (MHA_{BG0214}), de elevada especificidade, fácil implementação e de custos relativamente reduzidos, para a genotipagem dos subtipos e formas recombinantes dominantes (subtipo B, G, CRF02_AG e CRF14_BG) na epidemia do HIV-1 em Portugal. Por outro, pretendeu-se avaliar a aplicabilidade deste novo ensaio à genotipagem de um conjunto de estirpes circulantes entre indivíduos utilizadores de drogas injectáveis, onde fenómenos de recombinação são potenciados pelos comportamentos de risco associados.

2. Material e Métodos

2. Material e Métodos

2.1 Amostras de material biológico utilizadas neste estudo

2.1.1 Caracterização de amostras de referência

No decurso deste trabalho utilizaram-se dois tipos de amostras vulgarmente designadas do mesmo como “de referência”. Seis delas corresponderam a clones moleculares contendo genomas completos ou sequências genómicas parciais de HIV-1, clonadas em vectores plasmídicos. Em alternativa foram ainda usados extractos celulares totais (lisados) de uma linha celular cronicamente infectada com HIV-1. Estes foram previamente caracterizados como pertencentes aos subtipos B, G e à forma genética recombinante CRF02_AG. Como amostras de referência correspondentes a vírus do subtipo B foram utilizados um clone molecular infeccioso, contendo um genoma viral completo, designado pNL4-3 (AF324493, cedido pelo *AIDS Reserch and Reference Reagent Program*, NIH, EUA), assim como um extracto de uma linha celular linfoblastóide, de origem humana cronicamente infectada com HIV-1 designada 8E5LAV (38). Como representantes de sequências virais do subtipo G utilizaram-se 7 clones moleculares, todos eles derivados do vector plasmídico pGEM®-T Easy (Promega, EUA). Um destes clones, contendo um genoma viral completo, foi designado p2655 (AY612637), enquanto que os outros seis clones contendo genomas virais parciais, divididos em dois fragmentos (aproximadamente metade do genoma viral) sobreponíveis, foram designados 5'3037, 5'3306, 5'588, 3'3037, 3'3306, 3'588. Nestes, a metade 5' compreendia um fragmento de aproximadamente 4100pb (coordenadas 682-4776 em relação à estripe de referência HXB2), enquanto, que os clones com a metade 3' eram constituídos por um fragmento de aproximadamente 5100pb (coordenadas 4467-9537 em relação à estripe de referência HXB2). Estes clones genómicos foram previamente obtidos através da amplificação de DNA proviral por PCR de fragmentos longos (*Long-Range PCR*) preparado a partir de PBMC purificadas por centrifugação em gradiente de densidade (ver secção 2.2.1.2), obtidas de indivíduos habitantes da região metropolitana de Lisboa (A. Esteves, comunicação pessoal). Como amostra de

referência para a forma recombinante circulante CRF02_AG utilizou-se um clone molecular infeccioso, designado pBD6-15 (AY271690), obtido por clonagem, no vector PCR-XL-TOPO TA (Invitrogen, EUA), do genoma proviral do isolado primário 99CMBD6, amplificado por PCR (110). Este clone molecular foi gentilmente cedido pelo Dr. Hans-Georg Kräusslich (Universität Heidelberg, Alemanha).

2.1.2 Caracterização de amostras biológicas de origem clínica

Foram seleccionados para este trabalhos dois conjuntos de amostras biológicas provenientes de indivíduos infectados por HIV-1: plasma ou PBMC infectadas, conservadas nos laboratórios da Unidade de Ensino e Investigação de Virologia/IHMT (29, 30, 78), e as quais serviram como base para diversos estudos de caracterização genética de HIV-1. Todas estas amostras foram colhidas mediante consentimento informado, tendo sido acompanhadas do preenchimento (sempre que possível) de um inquérito epidemiológico que permitiu a obtenção de informação sobre uma série de parâmetros epidemiológicos básicos tais como a idade, género, nacionalidade, e comportamentos de risco (partilha de seringas para injeção de drogas por via endovenosa e relações sexuais não protegidas) dos indivíduos envolvidos neste estudo.

Um destes grupos incluiu um conjunto de 45 amostras provenientes de **UDI**, 30 das quais colhidas no Hospital Professor Doutor Fernando da Fonseca (Amadora-Sintra, AS) enquanto, que as restantes 15 foram colhidas no Centro de Atendimento a Toxicodependentes das Taipas em Lisboa (TA). Diferentes regiões do genoma proviral das estirpes de HIV-1 presentes nestas amostras foram previamente subtipadas por HMA e/ou sequenciação dos genes *gag*, e/ou *env* (29, 30, 78) e/ou *nef* (78), seguida da sua análise utilizando técnicas de inferência filogenética. Um segundo conjunto de amostras clínicas, num total de 5, foi também testado. Tais amostras, correspondendo a sangue total imobilizado em cartões FTA™ (Whatman/GE Healthcare, Alemanha) foram recolhidas no Centro de Saúde de Ponta Gea, na cidade da Beira, província de Sofala, em Moçambique. As repetições terminais longas (LTR), bem como porções dos genes *pol* e *env* das estirpes virais presentes neste grupo de amostras foram previamente subtipadas HMA e/ou sequenciação, seguida de análise filogenética (79).

As informações obtidas previamente para amostras biológicas foram determinantes na escolha das mesmas para utilização neste estudo. Os dados epidemiológicos das mesmas estão resumidos no anexo 1.

2.2 Preparação de amostras biológicas

2.2.1 Preparação de amostras de referência

O DNA correspondente aos clones moleculares de origem plasmídica contendo os genomas de HIV-1 e utilizados como referência, foram obtidos a partir de culturas bacterianas previamente transformadas com cada um deles. Para tal, foram obtidas suspensões saturadas (em meio LB, 94) de cada uma das estirpes bacterianas pretendidas, após a sua incubação a 37°C, com agitação, durante aproximadamente 12 a 16h. Posteriormente, estas mesmas suspensões foram submetidas a um processo de centrifugação a 6000g, durante 15 minutos a 4°C, desprezando-se o sobrenadante, enquanto, que o sedimento resultante foi conservado a -20°C até futura utilização. Por seu lado, para a obtenção de suspensões saturadas do clone molecular CRF02_AG, inicialmente fornecido como DNA plasmídico imobilizado em papel de filtro, este foi eluído em 100µl de solução tampão TE (10mM Tris-HCl, 1mM EDTA, pH8,0) por eluição passiva a 4°C durante a noite, a partir de um fragmento do mesmo contendo cerca de 5µg de DNA plasmídico. A suspensão de DNA assim obtida foi, em seguida, utilizada para transformar células competentes das estirpes TG1 (*supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5* (rK⁻ mK⁻) [F' *traD36 proAB lacI^f ZAM15*]) e XL1-Blue MRF' (*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* (F' *proAB lacI^f ZAM15*, Tn10 (Tetr)) (Stratagene, EUA) de *Escherichia coli*.

2.2.1.1 Preparação de células competentes de *Escherichia coli* e sua transformação por choque térmico

A preparação de células competentes de *E. coli* foi efectuada com base numa versão ligeiramente modificada de um protocolo inicialmente descrito na bibliografia (17). Resumidamente, as estirpes utilizadas foram inoculadas em meio LB a 37°C e mantidas com agitação durante a noite até à obtenção de suspensões saturadas. Em seguida 150µl destas foram utilizados para a inoculação de 15ml do mesmo meio de cultura, tendo as suspensões bacterianas diluídas, assim obtidas, sido incubadas nas condições de crescimento descritas em cima, até que as mesmas apresentassem uma densidade óptica variando entre 0,3-0,5 (leitura a 650nm). Em seguida, as células bacterianas foram recolhidas por centrifugação a 4500g, durante 10min, a 4°C, tendo sido desprezado o sobrenadante. O sedimento celular foi directamente ressuscitado em 500µl de TSS (do inglês, *Transformation and Storage Solution*) mantido a 4°C.

Para a transformação das células competentes, adicionaram-se 5µl de uma suspensão de DNA plasmídico a 200µl de células ressuscitadas em meio TSS, mantidas em microtubos previamente arrefecidos em gelo. As misturas foram homogeneizadas por agitação suave e em seguida incubadas, em gelo, durante 1h. Posteriormente, as células foram sujeitas a um choque térmico a 42°C durante 90seg, e colocadas, novamente, em gelo, por 5min. As suspensões bacterianas foram então adicionadas de 900µl de meio LB desprovido de antibiótico, e incubadas a 37°C durante uma hora, com agitação suave. Após esta incubação, 200µl das suspensões bacterianas foram semeados em caixas de Petri contendo meio LB suplementado com o antibiótico ampicilina (100µg/ml), X-Gal (80µg/ml) e IPTG (20µg/ml). As caixas foram mantidas numa estufa a 37°C durante aproximadamente 16h (durante a noite). Ainda que a presença de X-Gal e IPTG não fossem necessárias à identificação de colónias contendo o plasmídeo pBD6-15, a sua presença garantiu a possibilidade de identificação de eventuais colónias contendo plasmídeos deletantes, nos quais viesse a ser restabelecida a actividade da β -galactosidase como resultado de α -complementação.

2.2.1.2 Isolamento de plasma a partir de amostras de sangue periférico total

As amostras de plasma sanguíneo humano foram preparadas através da sua separação prévia das células sanguíneas por centrifugação (25000g, durante 10min) através de um gradiente de densidade (Lymphoprep™, Noruega). Para tal foram utilizadas amostras de 5-10ml de sangue periférico total, colhido para tubos devidamente identificados e contendo 1ml de agente anticoagulante (citrato de sódio a 3,8%, p/v). Uma vez separado, o plasma foi conservado a -80°C até à sua utilização.

2.2.1.3 Preparação das amostras de sangue imobilizado em cartões FTA™ (Whatman/GE Healthcare)

A colheita de sangue total destinado a ser imobilizado/conservado em cartões FTA™ foi feita por punção digital com lancetas descartáveis. Seguidamente, foram depositados cerca de 200µl de sangue total directamente sobre os cartões FTA os quais, uma vez secos, foram mantidos protegidos da humidade e à temperatura ambiente. Com a ajuda de um *Harris Micro-Punch* foram removidos dois discos de aproximadamente 1,2mm de diâmetro, os quais foram transferidos para microtubos (1,5ml). Entre cada uma das amostras processadas, foram efectuadas 5-7 punções na zona não utilizada do cartão FTA™, de modo a evitar possíveis contaminações.

Os discos foram submetidos a quatro lavagens com 400µl de *FTA™ Purification Reagent* (Whatman/GE Healthcare, Alemanha), seguidas de três lavagens com 300µl de solução tampão TE para eliminação de potenciais contaminantes inibidores da reacção de PCR. Entre cada lavagem efectuou-se um passo de incubação à temperatura ambiente durante 5min, com subsequente agitação dos discos num vórtex. Após secagem sob vácuo, os discos foram imediatamente utilizados como matriz em reacções de PCR (dois discos por reacção), ou conservados a 4°C até posterior utilização por um período máximo de uma semana.

2.3 Extracção de ácidos nucleicos

2.3.1 Extracção de ácidos nucleicos de amostras de referência

O DNA plasmídico dos diferentes clones moleculares contendo estirpes de HIV-1 de referência foi extraído a partir dos sedimentos celulares mantidos a -20°C (como referido em 2.1.1). A extracção do DNA plasmídico foi executada por recurso ao *kit* de extracção comercial *Plasmid Purification MID KIT* (QIAGEN, EUA) seguindo o protocolo recomendado pelo fabricante. As moléculas de DNA plasmídico assim obtidas foram ressuspensas em 100µl de H₂O e posteriormente submetidas a electrofore em gel de agarose a 1% em solução tampão TAE 0,5x (20mM Tris-Acetato, 0,5mM EDTA, pH 8.0) (94) na presença de 0,5µg/ml de brometo etídio. Para a visualização e gravação de imagens sob luz ultravioleta (UV) dos produtos extraídos nesta secção utilizou-se o sistema *Gel Doc XR System* (Bio Rad, EUA).

2.3.2 Extracção de RNA total a partir de plasma humano congelado a -80°C

A extracção de RNA a partir de amostras de plasma humano, mantidas a -80°C, foi realizada recorrendo ao *kit* de extracção rápida *INSTANT Virus RNA Kit* (*Analytikjena*, Alemanha). Este *kit* demonstrou uma boa eficiência na extracção de RNA total, tendo sido utilizado com sucesso em trabalhos realizados anteriormente (29, 30). O procedimento experimental foi realizado de acordo com as recomendações do fabricante. Brevemente, o protocolo de extracção pressupõe o descongelamento do plasma à temperatura ambiente, seguido de um passo de lise em tubos de extracção, lavagem das colunas onde se fixa o RNA e, por fim, a eluição do RNA extraído. A etapa de lise foi efectuada em laboratório de segurança biológica BSL-2 (BSL, do inglês *BioSafety Level*), numa câmara de fluxo laminar, onde foram adicionados 450µl de uma solução de lise (RL) a cada tubo de extracção. Posteriormente, a esse volume adicionou-se 150µl de plasma (manipulação efectuada em laboratório de nível de segurança biológica BSL-3). As misturas plasma/solução de lise foram homogeneizadas vigorosamente, 3 a 4 vezes, num vórtex, em pulsos de 10seg, durante um período de incubação de 15min à temperatura ambiente, seguida de centrifugação rápida. A

extracção prosseguiu com a adição de um volume de 600µl de solução tampão de ligação (*RBS*) às misturas anteriormente preparadas. Estas foram homogeneizadas por agitação forte (vórtex) e em seguida aplicadas nas colunas de extracção, previamente colocadas em tubos colectores. O processo foi iniciado com a adição às colunas de 650µl das misturas contendo o RNA, o qual a elas foi fixado durante uma breve centrifugação a 10000g durante 1min. O volume residual de cada amostra foi então adicionado à mesma coluna e o processo repetido. Depois de remover possíveis contaminantes através da adição sucessiva das soluções de lavagem *HS* seguida de *LS*, e pela centrifugação dos tubos de extracção a 10000 g durante 1min, o RNA foi eluído pela adição de 30µl de água livre de RNases. Após incubação (2min) à temperatura ambiente, o volume eluído foi recolhido por centrifugação (1min a 10000g) e armazenado, à temperatura de -80°C, em duas alíquotas de 5µl e uma de 20µl, até futura utilização.

2.4 Quantificação espectrofotométrica de amostras de DNA

Os extractos de DNA plasmídico foram quantificados num espectrófotometro Smart SpecTM 3000 (Bio-Rad, EUA) através da análise da absorvância, a 260 nm, de diluições preparadas a partir dos extractos concentrados. De igual modo, e quando necessário, os fragmentos de DNA amplificados através de algumas das reacções de PCR (secção 2.8.1) foram diluídos em água Milli-Q na razão de 1:140 para, em seguida, as absorvâncias destas soluções serem lidas a 260nm, permitindo a quantificação de DNA nas mesmas. O grau de pureza das amostras foi ainda avaliado através da razão entre as absorvâncias registadas a 260 e 280nm. Após registo das concentrações foram efectuadas as diluições necessárias de modo a que as matrizes utilizadas posteriormente na secção 2.8.4 apresentassem uma concentração semelhante por cada região analisada.

2.5 Síntese, por reacção de transcrição reversa, de DNA complementar (cDNA) a partir de RNA total

Para a síntese de cDNA a partir das moléculas de RNA, extraídas como descrito na secção anterior, foi utilizado o *kit* comercial *RevertAidTM H Minus First Strand cDNA Synthesis Kit* (Fermentas, Lituânia). O protocolo experimental foi executado de acordo com as instruções recomendadas pelo fabricante. Genericamente, a 5µl de extracto de RNA foram adicionados 1µl de oligonucleótidos (hexâmeros) sintéticos iniciadores com sequência aleatória, e 6µl de água tratada com DEPC (Diethyl pirocarbonato). A mistura foi incubada a 70°C durante 5min (desnaturação do RNA) e, em seguida, colocada em gelo durante igual período de tempo. Seguidamente, foi preparada uma mistura de reacção contendo 4µl solução tampão de reacção 5X, 1µl inibidor de RNAses (*RiboLockTM RNase Inhibitor*), 2µl dNTPs (10mM) (concentração final, 1mM), 1µl de transcriptase reversa (*RevertAidTM H Minus M-MuLV*, concentrada a 200U/µl), MgCl₂, (6mM). A preparação das misturas de reacção foi realizada em câmara de fluxo laminar durante a qual estas foram mantidas no gelo. As reacções de transcrição reversa foram efectuadas num termociclador iCycler de acordo com o programa representado na Tabela 2.1. A reacção de transcrição origina um produto final de cDNA a partir das moléculas de RNA extraído. Os produtos resultantes da transcrição reversa (cDNA) foram conservados a -20°C até utilização futura.

Tabela 2.1 - Condições utilizadas para as reacções de transcrição reversa a partir de RNA total.

	Temperatura (°C)	Tempo	Número de ciclos
Hibridação	25	5min	1
Extensão	42	60min	1
Inactivação da enzima RT	72	15min	1

2.6 Construção de alinhamentos múltiplo de sequências de HIV-1 para delineamento de *primers*/sondas

Foi construído um alinhamento múltiplo de 57 sequências correspondendo a genomas completos de estirpes de HIV-1 disponíveis na base de dados de acesso livre *Los Alamos sequence database* (acessível em <http://hiv-web.lanl.gov>). No alinhamento múltiplo em questão foram incluídas 19 sequências correspondentes a genomas do subtipo B, 12 ao subtipo G, 2 para cada um dos subtipos A e F, 1 para cada um dos subtipos K, C, J, H e 18 correspondentes a sequências genómicas das formas recombinantes circulantes CRF02_AG (n=11) e CRF14_BG. (n=7). O alinhamento múltiplo foi efectuado e analisado, respectivamente, com recurso aos programas informáticos *MAFFT version 6* (disponível em <http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>) (52) e *BioEdit version 7.0.9.0* (disponível em <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) (43).

2.7 Delineamento de *primers* para PCR e de sondas para PCR em tempo-real

Os *primers*/sondas utilizados nas amplificações universais e específicas (secções 2.8.2 a 2.8.6 e 2.9) que permitiram, recorrendo a abordagens de tipo *nested*-PCR ou PCR em tempo-real, a amplificação de 8 regiões do genoma de HIV-1 (sequências codificantes das proteínas Gag, Protease, Transcriptase Reversa, Integrase, Vpr, Rev, Gp120 e Gp41) foram desenhados, manualmente, após análise do alinhamento referido na secção anterior. De acordo com a formulação convencional de ensaios de tipo MHA foi dada especial atenção à identificação de regiões do genoma viral cuja sequência se demonstrou ser distinta em cada um dos subtipos ou formas recombinantes circulantes em análise. Por seu lado, o desenho dos *primers* para amplificação universais de fragmentos do genoma de HIV-1 foi determinado pela identificação de regiões conservadas ao longo do genoma viral, situadas a montante e a jusante das regiões anteriormente mencionadas.

Os *primers*/sondas foram desenhados de forma, a que, na medida do possível, simultaneamente apresentassem um conteúdo C/G a variar entre os 30 e 60%, tamanho entre 22-25 nt e que as respectivas temperaturas de hibridação não ultrapassassem os

65°C onde o intervalo de temperaturas ideal se situa entre os 52-58°C. Por seu lado a presença de pelo menos 3 bases nucleotídicas (C/G) nas extremidades habitualmente confere uma maior especificidade devido ao número de pontes de hidrogénio estabelecidas, contudo a presença de um maior número de bases deve ser evitada (92, 97). No caso dos *primers* específicos/sondas contemplaram-se regiões alvo que apresentassem o maior número possível de nucleótidos diferentes quando as sequências de vírus dos vários subtipos foram comparadas entre si. A sequência/região alvo do genoma viral, orientação (sentido directo ou reverso) e coordenadas dos *primers*/sondas utilizadas neste trabalho encontram-se listados no anexo 2.

2.8 Amplificações de fragmentos de DNA por reacção em cadeia da Polimerase (PCR)

2.8.1 Verificação, por PCR, da integridade estrutural das amostras de referência (clones plasmídicos de genoma total ou parcial)

A integridade do DNA plasmídico referente às amostras de referência foi testada através da amplificação, por PCR, de cinco regiões distintas do genoma do HIV-1 [genes *gag*, *pol* (RT), *env* (gp120-região C2V3C3, e ectodomínio de gp41) e *nef*]. Para tal, recorreu-se a combinações de *primers* específicos para cada uma das regiões, os quais se demonstraram, em estudos anteriores (29, 30, 78, 104), suficientemente específicos para permitirem a amplificação indubitável de cada uma das regiões pretendidas. As sequências, regiões “alvo”, orientação e coordenadas dos *primers* utilizadas nestas amplificações preliminares encontram-se identificados no anexo 2, onde são representados pela designação abreviada.

As mistura de reacção, preparadas numa câmara de fluxo laminar para um volume final de 50µl, incluíram 25µl de uma de PCR pronta a usar (*PCR Master Mix* 2x, Fermentas, Lituânia), contendo 0,05U/µl de Taq DNA Polimerase em solução tampão de reacção, 4mM MgCl₂, 0,4mM de cada um dos 4 dNTPs, 19µl de água Milli-Q, 1µl de cada um dos respectivos *primers* (concentrados a 10pmol/µl, correspondendo a 200nM final em cada uma das reacções de amplificação) e 2µl dos respectivos DNA alvo a amplificar numa concentração final que variou entre os 8,4x10⁻³ng/µl e 2,6x10⁻¹

ng/μl (diluições de 1:100). Para os casos em que não se visualizaram produtos amplificados em gel de agarose repetiram-se as amplificações partindo de concentrações superiores. As condições da reacção de PCR utilizadas estão representadas na Tabela 2.2. Os produtos amplificados foram visualizadas após terem sido separados por electroforese em gel de agarose a 1,5% em solução tampão TAE 0,5x na presença de 0,5μg/ml de brometo etídio. Os respectivos tamanhos foram estimados a partir da migração relativa de um marcador de massa molecular contendo de moléculas lineares de DNA em cadeia dupla com tamanhos a variar de 100pb (*GeneRuler™ 100 bp DNA Ladder Plus*, Fermentas, Lituânia). Para a visualização e registo em suporte digital de imagens sob luz ultravioleta (UV) dos produtos amplificados (nesta secção e em geral ao longo deste trabalho) utilizou-se o sistema *Gel Doc XR System* (Bio-Rad, EUA). A preparação das misturas de reacção e as adições de DNA às mesmas foram efectuados em compartimentos separados. Em todas a reacções de PCR foram contemplados controlos negativos de modo a avaliar possíveis contaminações durante o processo de amplificação. As reacções de PCR decorreram num termociclador *iCycler* (Bio-Rad, EUA).

Tabela 2.2 - Condições utilizadas para a amplificação por PCR dos fragmentos de DNA para as cinco regiões do genoma das estirpes virais de referência (*gag*, *pol* (RT), *env* (C2V3C3 e gp41), e *nef*).

	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	94	60seg	1
Desnaturação	94	45seg	35
Hibridação	50	45seg	
Extensão	72	120seg	
Extensão final	72	5min	1

2.8.2 Amplificação de fragmentos de DNA (universais e subtipo-específicos) para as várias regiões em estudo por PCR: amostras de referência

Para a amplificação de sequências genómicas de HIV-1 sob a forma de fragmentos de DNA designados de externos e internos foram utilizados *primers* universais, ao passo que a utilização de *primers* subtipo-específicos foi reservada para a terceira das amplificações efectuadas. Genericamente, os fragmentos de DNA foram amplificados numa primeira reacção universal, sendo utilizados como matriz para uma segunda e o produto da segunda utilizado numa terceira reacção de amplificação específica. Os *primers* utilizados na primeira reacção (externos) ligam-se numa região que limita exteriormente as sequências alvo dos *primers* utilizados na segunda reacção (internos). Para a terceira reacção utilizou-se um *primer* comum ao da reacção que a antecede, juntamente com um *primer* específico para o a testar subtipo (o esquema do protocolo utilizado encontra-se representado na Figura. 2.1).

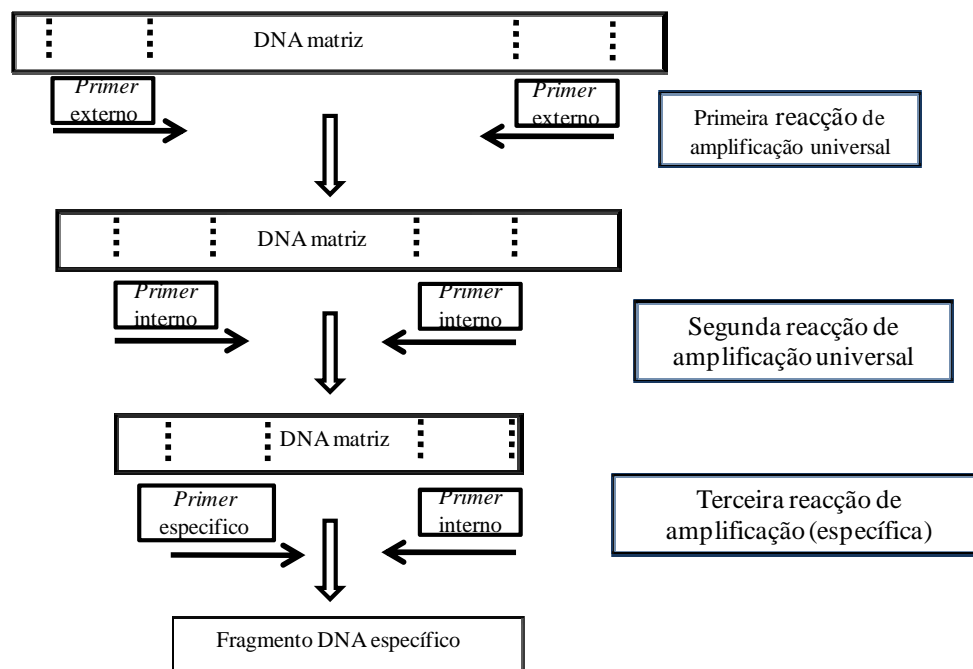


Figura 2.1 – Esquema representativo da estratégia amplificação utilizada na amplificação universal e específica dos fragmentos de DNA para as várias regiões do genoma do HIV-1.

As misturas de reacção para as amplificações por PCR dos fragmentos universais externos foram preparadas para um volume final de 25µl, que incluiu 12,5µl de uma pré-mistura para PCR (*PCR Master Mix 2x*, Fermentas, Lituânia), 9,5µl de água Milli-Q, 1µl de cada um dos respectivos *primers* (diluídos a 10pmol/µl, correspondendo a 400nM final em cada uma das reacções de amplificação) e 1µl das respectivas suspensões de DNA alvo a amplificar nas concentrações referidas na secção anterior (ver secção 2.8.1). Para a segunda reacção de amplificação (fragmento universal interno) e terceira reacção (amplificações específicas) as condições foram semelhantes às indicadas acima. No entanto, utilizou-se como DNA matriz 1µl do produto amplificado na reacção antecedente. Os produtos amplificados foram diluídos na razão de 1:1000 (sempre que fosse visível em gel de agarose uma banda correspondente ao fragmento esperado) entre reacções consecutivas, de modo a que um excesso de matriz original não viesse a inibir a reacção subsequente ou, de outra forma, se viessem a obter produtos inespecíficos. As condições de PCR utilizadas para as amplificações universais encontram-se representadas na Tabela 2.3. Para a amplificação dos fragmentos específicos, foram previamente ensaiadas diferentes de temperaturas de hibridação (55, 58, 60, 63, 66 e 69°C), através do estabelecimento de um gradiente de temperatura no bloco do termociclador, com o objectivo de definir quais as condições que se demonstrassem mais restritivas para cada uma das amplificações desejadas. As condições da reacção de amplificação encontram-se representadas na Tabela 2.4.

Os produtos amplificados foram visualizados sob luz UV, após migração em gel de 1,5% agarose (ver secção 2.3.1). Para a análise de fragmentos de massa molecular inferior a 100pb, a concentração de agarose foi aumentada para 2% de modo a que se obtivesse uma melhor resolução dos mesmos. Em alternativa, a análise de pequenos fragmentos de DNA foi conseguida através da sua separação electroforética em gel de poliacrilamida (8%) (94), tendo a sua massa sido estimada por comparação com a migração de um marcador de massas moleculares apropriado (*pUC Mix Marker*, Fermentas, Lituânia). Uma vez terminada a electroforese, os géis de acrilamida foram imersos numa solução de brometo de etídeo (0,5 µg/ml), tendo as moléculas de DNA sido reveladas sob luz UV. Os *primers* utilizados neste estudo foram obtidos comercialmente a partir das empresas Bonsai Technologies, STAB VIDA ou Alfagene.

Tabela 2.3 - Condições utilizadas para amplificação, por PCR, dos fragmentos externos e internos para as várias regiões do genoma do HIV em estudo.

	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	94	60seg	1
Desnaturação	94	45seg	35
Hibridação	55	45seg	
Extensão	72	60seg	
Extensão final	72	5min	1

Tabela 2.4 - Condições utilizadas para amplificação dos fragmentos específicos para as várias regiões do genoma do HIV em estudo por reacção de PCR.

	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	94	60seg	1
Desnaturação	94	45seg	35
Hibridação	55, 58, 60, 63, 66, 69	45seg	
Extensão	72	60seg	
Extensão final	72	5min	1

2.8.3 Amplificação de sequências virais por PCR a partir de amostras biológicas

Para a amplificação dos fragmentos do genoma de HIV-1 a partir de amostras clínicas foi utilizado o protocolo otimizado, na secção anterior (ver secção 2.8.2), para as amostras de referência. No entanto, e de forma a aumentar a sensibilidade e especificidade das amplificações obtidas, recorreu-se ao sistema *puRe Taq™ ready-To-Go™ PCR Beads* (GE Healthcare, Alemanha), tendo as reacções de PCR sido preparadas segundo as indicações do fabricante. Este tipo de sistema é apresentado comercialmente sob a forma de microtubos em que os reagentes necessários para a reacção de amplificação se apresentam liofilizados numa esfera. As reacções de PCR foram preparadas para um volume de 25µl final contendo os seguintes reagentes nas seguintes concentrações finais: 10mM Tris-HCl (pH 9.0), 200µM cada dNTP, 50mM KCl e 1,5mM MgCl₂. Cada uma destas misturas de reacção contém ainda 2,5 unidades de uma polimerase de DNA (*PuReTaq DNA polymerase*).

O procedimento de amplificação foi efectuado em duas variantes. Na primeira, correspondendo à primeira reacção de amplificação universal, cada um dos tubos foi reconstituído para um volume de final de 25µl com 16µl de água Milli-Q, 1µl de cada um dos respectivos *primers* universais (a partir de uma solução inicial 10pmol/µl, correspondendo a 400nM final) e 5µl dos respectivos DNA alvo a amplificar (produtos resultantes da reacção de transcrição reversa). Finda a primeira reacção de amplificação, 5µl do produto amplificado foram transferidos para um novo tubo e sujeitos à segunda reacção amplificação com *primers* universais. Nas situações em que a quantidades de produto amplificado resultante da primeira reacção foram elevadas procedeu-se a uma diluição (1:1000) em água e utilizado 1µl de produto para iniciar a segunda reacção. A terceira reacção de amplificação específica foi iniciada com 1µl (razão de 1:1000) dos produtos amplificados na reacção anterior, na qual seriam utilizados em conjunto um *primer* de amplificação específico e outro de amplificação universal.

Na segunda variante, o todo o procedimento foi semelhante ao descrito em cima, tendo-se reduzido apenas o número de reacções, uma vez que a reacção de amplificação específica foi iniciada directamente com 1µl de produto de amplificação, na razão de 1:1000, resultante da primeira reacção universal (quando as quantidades de produto amplificado foram elevadas). Os produtos amplificados foram visualizados sob luz UV, após migração em gel de 1,5% agarose, tal como descrito anteriormente (ver secção 2.8.1).

As condições iniciais das reacções de amplificação foram semelhantes às experimentadas para as amostras de referência na secção anterior (ver secção 2.8.2). Contudo, ao longo do protocolo, sempre que se considerou necessário procedeu-se a ajustes na temperatura de hibridação (ver capítulo 3 secção 3.4.2).

2.8.4 Amplificação dos fragmentos de DNA por PCR em tempo-real: ensaio de competição de matrizes (amostras de referência)

A amplificação específica, por PCR, de fragmentos do genoma de HIV-1 em tempo-real foi testada em ensaios de competição. Estes foram realizados (seguindo uma abordagem de PCR em tempo-real) na presença de *SYBR*[®] *Green* I (intercalador fluorescente de moléculas de DNA em cadeia dupla) e foram delineados para que numa

mesma mistura de reacção fosse contemplada a presença de duas matrizes de DNA. Uma delas, designada de DNA competidor, correspondeu a um fragmento de DNA comportando sequências virais que, embora colineares com as que se pretendiam obter (DNA alvo, ver abaixo), não deveriam, idealmente, ser amplificadas de forma específica pelo par de *primers* utilizado. Em contrapartida, a outra matriz correspondeu a DNA dito alvo, cuja amplificação específica se pretendia detectar. Foram assim preparadas várias misturas de reacção, nas quais a concentração do competidor foi mantida constante, ao passo que a concentração da amostra alvo utilizada variou, para que a razão entre ambas (competidor/alvo) em cada uma das misturas de reacção obedecesse às razões aproximadas de 1:1, 10:1 e 100:1. Esta abordagem baseia-se no facto de que a fluorescência emitida pelas moléculas é proporcional à quantidade de produto amplificado resultante uma vez que o *SYBR® Green I* apresenta a particularidade de ser um intercalador fluorescente de DNA em cadeia dupla (57) (Figura 2.2).

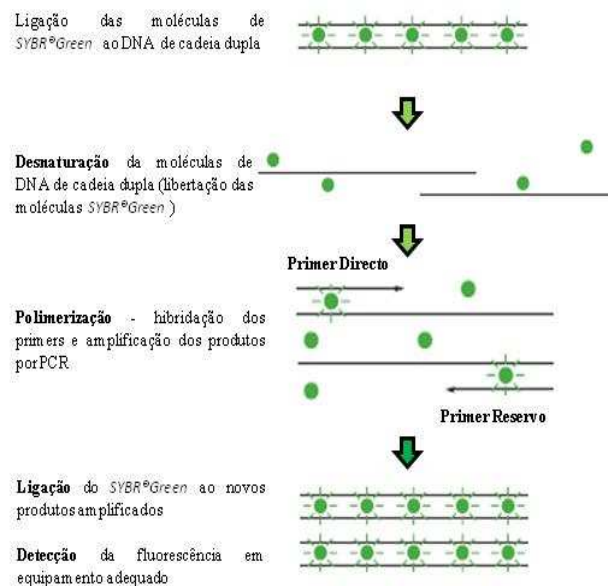


Figura 2.2 – Representação esquemática dos principais passos utilizados no sistema de PCR em tempo-real em que a detecção dos fragmentos de DNA amplificados é feita como consequência da presença de *SYBR® Green I* no meio de reacção (adaptado, www3.appliedbiosystems.com).

Esta estratégia permitiu a detecção de um aumento de sinal fluorescente, atingindo este uma fase exponencial a partir de um ponto (C_t) que é função da concentração de DNA alvo presente na mistura de amplificação. A mistura de PCR pronta a usar utilizada incluía, ainda, a presença de uma polimerase de DNA termorresistente, e quimicamente modificada, que se apresentava inactiva à temperatura ambiente (mas passível de ser activada pelo calor), impedindo que hibridações não específicas dos *primers* pudessem iniciar uma extensão inespecíficas. Esta característica conferiu uma maior especificidade às amplificações de DNA pretendidas. Finalmente, a presença de KCl e $(\text{NH}_4)_2\text{SO}_4$ na solução tampão de reacção garantiu uma elevada especificidade na hibridação dos *primers*. A existência de um controlo interno (*ROX*) permitiu a normalização do registo de fluorescência emitida pelo *SYBR[®] Green I*, corrigindo eventuais variações de pipetagem ou de detecção do sinal de fluorescência pelo aparelho utilizado.

As misturas de amplificação de fragmentos de genoma viral, por PCR em tempo-real, em que se utilizou um *primer* de amplificação específica e um dos *primers* internos universais (ver anexo 2) foram preparadas para um volume final de 12,5µl utilizando 6,25µl de uma pré-mistura de PCR (*Maxima[™] SYBR[®] Green I qPCR Master Mix 2X*, Fermentas, Lituânia; ver capítulo 3 secção 3.3.4), 0,25µl de água Milli-Q, 1µl de cada um dos respectivos *primers* (concentrados a 10pmol/µl, correspondendo a uma concentração final de 800nM em cada uma das reacções de amplificação) e 2µl de cada um dos diferentes DNA alvo e competidor a utilizar. Como referido anteriormente na secção 2.4 as concentrações dos DNA matriz foram equilibradas para a mesma região analisadas (ver capítulo 3 secção 3.1.2). No caso do controlo negativo do ensaio, o DNA foi substituídos por 4µl água Milli-Q. Em todas as reacções foi incluído ainda um controlo positivo contendo 2µl DNA alvo e 2µl de água Milli-Q (desprovido de DNA competidor). As condições da reacção foram semelhantes para todos os ensaios, variando apenas a temperatura de hibridação consoante o par de *primers* utilizados. As condições de reacção utilizadas nestes ensaios para cada região por PCR em tempo-real encontram-se representadas na Tabela 2.5.

Todas as misturas de reacções foram preparadas em microtubos de 0,2ml (*PCR-02-C, thin wall flat cap*, Axygen, EUA) e todos os ensaios de amplificação por PCR em tempo-real decorreram num termociclador *Research Rotor-Gene 3000* (Corbett

Research/QIAGEN, EUA), tendo a monitorização da amplificação sido avaliada através do programa informático *Rotor-gene 6 version 6.1* (Corbett Research/QIAGEN, EUA).

Tabela 2.5 - Condições utilizadas para amplificação de secções do genoma de HIV-1 por PCR em tempo-real.

	Região alvo do genoma viral	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	Todas as regiões	94	60seg	1
Desnaturação	Todas as regiões	94	45seg	
Hibridação	Gag	69	45seg	35
	Protease, Integrase, Rev, Gp120	58		
	RT	60		
	Gp41	63		
Extensão	Todas as regiões	72	60seg	
Extensão final	Todas as regiões	72	5min	1

2.8.5 Detecção de amplificação específica de sequências de HIV-1 (amostras biológicas), por PCR em tempo-real realizado na presença de *SYBR[®] Green I*

Para a amplificação, por PCR em tempo-real e a partir de amostras clínicas, de fragmentos do genoma de HIV-1 usando *primers* específicos, foi utilizado um protocolo no qual os produtos resultantes da primeira, ou segunda reacção de amplificação mencionada anteriormente (ver secção 2.8.2) serviram de molde para a reacção de amplificação específica. Esta amplificação foi detectada, em tempo-real devido à presença de *SYBR[®] Green I* nas misturas de reacção. Para estas reacções foram incluídos um *primer* específico e outro universal/interno.

As misturas de reacção foram preparadas para um volume final de 12,5µl que incluiu 6,25µl de uma mistura pronta a usar de PCR (*Maxima[™] SYBR[®] Green I qPCR*

Master Mix 2X, Fermentas, Lituânia), 3,25µl de água Milli-Q, 1µl de cada um dos respectivos *primers* (diluídos a 10pmol/µl, correspondendo a 800nM final em cada uma das reacções de amplificação) e 1µl de cada um dos respectivos DNA alvo. Na preparação do controlo negativo da reacção de amplificação, o DNA matricial foi substituído por 1µl água Milli-Q. Nas situações onde não se observou o produto resultante da primeira reacção universal de PCR, aumentou-se o volume de DNA matricial para 2,5µl. As condições da reacção foram semelhantes para todos os ensaios, variando apenas a temperatura de hibridação consoante o par de *primers*/região pretendida de acordo com o procedimento experimentado anteriormente (ver secção 2.8.3).

2.8.6 Amplificação específica de sequências virais (Gp41) de HIV-1 (amostras clínicas) por PCR em tempo-real usando sondas de hidrólise (*TaqMan*)

A detecção de amplificação específica, por PCR em tempo-real, de parte da sequência codificante da glicoproteína Gp41 de HIV-1, foi ainda levada a cabo por recurso a sondas do tipo *TaqMan*. Nestes casos foram utilizados um par de *primers* internos em conjunto uma sonda *TaqMan* específica para a região/subtipo a detectar. Nestes ensaios foram ainda utilizados como DNA matricial o resultado da primeira, ou da segunda reacção de amplificação universal, tal como descrito nos passos anteriores (ver secção 2.8.2).

As misturas de reacção foram preparadas para um volume final de 12,5µl que incluiu 6,25µl de uma mistura de PCR pronta a utilizar (*DyNAmo™ Flash Probe qPCR 2X*, Finnzymes, Finlândia), 2,25µl de água Milli-Q, 1µl de cada um dos *primers* a partir de uma solução inicial a 10pmol/µl, (800nM final em cada uma das reacções de amplificação) e 1µl de sonda *TaqMan* partir de uma solução inicial a 3pmol/µl (240nM final) e 1µl dos respectivos DNA matricial. No caso do negativo de reacção o DNA matricial foi substituído por 1µl água Milli-Q. Quando se utilizou como matriz o DNA amplificado na primeira reacção universal ajustaram-se os volumes da mistura de reacção aumentando o volume de DNA matricial para 2,5µl.

Os ensaios de amplificação específica foram ainda efectuados na presença de duas sondas *TaqMan* em simultâneo numa mesma mistura de reacção (abordagem

multiplex). Estas foram intencionalmente marcadas com fluorocromos excitáveis e emissores de radiação de comprimento de onda distintos (JOE 548nm e Rox 591-604nm). As condições da reacção de amplificação foram semelhantes para todos os ensaios (Tabela 2.6). As sequências dos *primers* e sondas encontram-se descritos no anexo 2.

Tabela 2.6 - Condições utilizadas para a detecção específica de parte da sequência codificante de Gp41 por PCR em tempo-real usando sondas *TaqMan*.

Detecção a 3 passos	Temperatura (°C)	Tempo	Número de ciclos
Desnaturação inicial	94	60seg	1
Desnaturação	94	45seg	35
Hibridação	68	45seg	
Extensão	72	45seg	
Extensão final	72	5min	1

2.9 Amplificação, purificação e sequenciação de produtos de PCR

A análise da sequência nucleotídica de alguns dos fragmentos de PCR amplificados foi efectuada com base na sequenciação de amplicões obtidos para cada uma das regiões do genoma em estudo. Para tal, foi utilizado um protocolo de *heminested*-PCR usando como DNA matricial os produtos resultantes da primeira reacção com *primers* universais (ver Figura 2.1, secção 2.8.2). Esta abordagem (*heminested*-PCR) foi necessária uma vez que os produtos específicos resultantes da terceira reacção de amplificação não apresentavam tamanho suficiente para uma avaliação satisfatória da sua identidade genética por análise filogenética. Para estas reacções foram utilizados uma combinação de *primers* composta por um dos *primers* universais utilizados na primeira reacção, para além de um segundo, normalmente utilizado nas segundas reacções de amplificação.

As misturas de reacção foram preparadas para um volume final de 50µl que incluiu 25µl de uma mistura *PCR* pronta a utilizar *Master Mix 2x* (Fermentas. Lituânia), 19µl de água Milli-Q, 2µl de cada um dos respectivos *primers* a partir de uma solução

inicial a 10pmol/ μ l, (400nM final) e 2 μ l dos respectivos DNA alvo a amplificar (produtos amplificados dos fragmentos externos). As condições da reacção de PCR utilizadas foram idênticas às utilizadas às descritas na secção 2.8.2 para as amplificações universais.

Os produtos amplificados foram purificados recorrendo a um *kit* comercial (Wizard®PCR Preps DNA Purification System, Promega, EUA) segundo as instruções fornecidas pelo fabricante. Em termos gerais, neste tipo de sistema o DNA é fixo a uma coluna por adição prévia de uma resina, seguindo-se uma lavagem com isopropanol a 80% para remoção de moléculas adsorvidas à resina de forma inespecífica. A libertação do DNA da coluna é conduzida pela adição de 50 μ l de água tratada com DEPC. Posteriormente, 3 μ l de produto de purificação foram analisados por electroforese em gel 1,5% agarose (ver secção 2.8.1), avaliando assim o rendimento aproximado do processo de extracção e o grau de purificação obtido.

Os produtos purificados foram sequenciados pelos serviços de sequenciação da empresa STAB VIDA (Portugal), recorrendo a uma variação da técnica de terminação da polimerização em cadeia originalmente descrita por Sanger (95). A sequenciação foi efectuada apenas num dos sentidos usando um dos *primers* da amplificação anterior (o *primer* que apresenta-se menor números de bases degeneradas) para cada região do genoma viral analisada.

2.10 Análise filogenética das sequências nucleotídicas

As sequências nucleotídicas resultantes da sequenciação directa de cada um dos produtos de amplificação foram alinhadas juntamente com sequências de referência de de HIV-1 dos diferentes grupos, subtipos e sub-subtipos, disponíveis na base de dados de acesso livre *Los Alamos sequence database*. Para a construção, e edição, dos alinhamentos múltiplos, foram utilizados os programas informáticos *MAFFT version 6* e *BioEdit version 7.0.9.0*, tal como mencionado anteriormente.

As análises de inferência filogenética foram realizadas através da construção de árvores filogenéticas com o programa Mega vs. 4 (108) As árvores filogenéticas foram obtidas pelo método de junção de vizinhos (*Neighbor-Joining*) (93) a partir de matrizes de distâncias genéticas calculadas com base nas distâncias corrigidas através da fórmula

de Kimura a dois parâmetros (K2P) (58). Foi ainda avaliada a robustez das topologias sugeridas pela execução de testes de *bootstrapping* (31), com base em 1000 amostragens aleatórias dos dados originais. Para os agrupamentos que apresentaram valores de *bootstrap* iguais ou superiores a 80% consideram-se agrupamentos consistentes do ponto de vista filogenético.

2.11 Análise de recombinantes

As amostras biológicas para as quais a atribuição de um genótipo às sequências analisadas revelou resultados controversos foram ainda analisadas por *BootScanning* de modo a avaliar a sua possível origem recombinante. Para o efeito recorreu-se ao programa bioinformático *Simplot* (64) utilizando o mesmo conjunto de sequências de referência da secção anterior (ver secção 2.10). Esta análise foi efectuada considerando árvores filogenéticas de NJ (93) com distâncias genéticas baseadas no modelo de K2P (58), estimadas numa janela de análise de 180pb, deslocando-se ao longo da sequência a analisar com incrementos de 10pb, e 1000 replicados (teste de *bootstrapping*). Nesta análise apenas foram considerados como irrefutáveis os resultados em que a associação da sequência em análise a um genótipo determinado apresentassem valores de *bootstrap* (31) iguais ou superiores a 70%.

3. Resultados

3. Resultados

3.1 Preparação das amostras de referência: extracção e quantificação de ácidos nucleicos

O DNA plasmídico dos diferentes clones moleculares contendo estirpes de HIV-1 de referência foram extraídos a partir de culturas saturadas de *Escherichia coli* (tal como descrito no capítulo 2 secção 2.3.1, tendo o sedimento de DNA sido ressuspendido num volume final de 100µl de H₂O. Após análise electroforética, em gel de agarose a 1%, de 3µl da suspensão obtida foi possível confirmar a presença, em todas elas, de moléculas de DNA plasmídico.

Os extractos de DNA plasmídico foram quantificados com exactidão após fracções destes terem sido diluídos em água Milli-Q na razão de 1:140. Esta quantificação foi efectuada através do registo da absorvâncias (a 260nm) destas soluções. O extracto de DNA (lisado celular) correspondente à amostra 8E5LAV foi igualmente quantificada pelo o mesmo método. As concentrações obtidas na suspensão inicial variaram entre os 14 e os 665 µg/ml. Contudo, antes da amplificação dos fragmentos virais foram efectuadas várias diluições (1:5, 1:10, 1:20, 1:100) da suspensão inicial sendo descrito nas secções seguintes o intervalo de concentrações finais utilizadas (ver secção 3.3.1).

De um modo semelhante foram ainda quantificados com exactidão os produtos resultantes das amplificações dos fragmentos de DNA obtidos mediante a utilização dos diferentes pares de *primers* externos (ver secção 3.3.2) usando, como matriz, as amostras de referência pNL4-3 (subtipo B), pBD6-15 (CRF02_AG) e p2655 (subtipo G). Estes fragmentos de DNA foram utilizados nos ensaios de competição entre matrizes, realizados recorrendo ao PCR em tempo-real (ver secção 3.3.4). Assim sendo, e tendo em conta os resultados obtidos pela leitura de absorvâncias, efectuaram-se as diluições necessárias de modo a que, por cada região analisada, as diferentes amostras DNA matriz utilizadas nas reacções de amplificação apresentassem, aproximadamente, a mesma concentração (Tabela 3.1).

Tabela 3.1 – Concentração do DNA correspondente aos fragmentos externos para as amostras de referência (pNL4-3 subtipo B, p2655 subtipo G e CRF02_AG).. Estão representados os valores de concentração equilibradas entre as amostras de referência para cada região e a concentração inicial utilizada nos ensaios de competição (diluição 1:1000). A concentração de DNA foi calculada através da leitura das absorvâncias a 260nm.

Região	Concentração final diluição(ng/μl) (1:1000)
Gag	$5,12 \times 10^{-2}$
Pr	$3,84 \times 10^{-2}$
RT	$5,76 \times 10^{-2}$
Rev/ Gp120	$4,8 \times 10^{-2}$
Gp41	4×10^{-2}

3.2 Delineamento de *primers*/sondas para PCR a partir de alinhamentos múltiplos de sequências de HIV-1

O desenho dos oligonucleótidos usados, no decurso deste trabalho, como *primers* ou sondas, foi efectuado por base num alinhamento múltiplo de sequências de estirpes de referência de HIV-1. O alinhamento em questão incluiu 57 sequências correspondendo a genomas completos de estirpes de HIV-1 disponíveis na base de dados *Los Alamos sequence database* (acessível em <http://hiv-web.lanl.gov>). Neste alinhamento múltiplo incluíram-se 19 sequências correspondentes a genomas de vírus do subtipo B, 12 a vírus do subtipo G; duas correspondentes a vírus incluídos em cada um dos subtipos A e F; uma para cada um dos subtipos K, C, J, H; 18 corresponderam a sequências genómicas das formas recombinantes circulantes CRF02_AG (n=11) e CRF14_BG (n=7). A escolha deste conjunto de sequências baseou-se no facto destes serem os subtipos e CRF mais frequentes em Portugal (30, 75, 76, 78).

A partir da análise visual do alinhamento referido em cima, foi possível identificar 7 regiões distribuídas ao longo do genoma [correspondendo estas a secções das regiões codificantes das proteínas Gag, Protease (Pr), Transcriptase Reversa (RT), Integrase (In), Rev, Gp120 e Gp41] altamente conservadas na sua totalidade (ou quase totalidade) quando as sequências de vírus de um mesmo subtipo/CRF eram comparadas entre si. No entanto, estas mesmas regiões apresentaram-se distintas quando, em alternativa, se comparavam sequências de vírus classificados em subtipos/CRF distintos. Estas regiões foram utilizadas para o delineamento dos *primers*/sondas específicas (Figura 3.1).

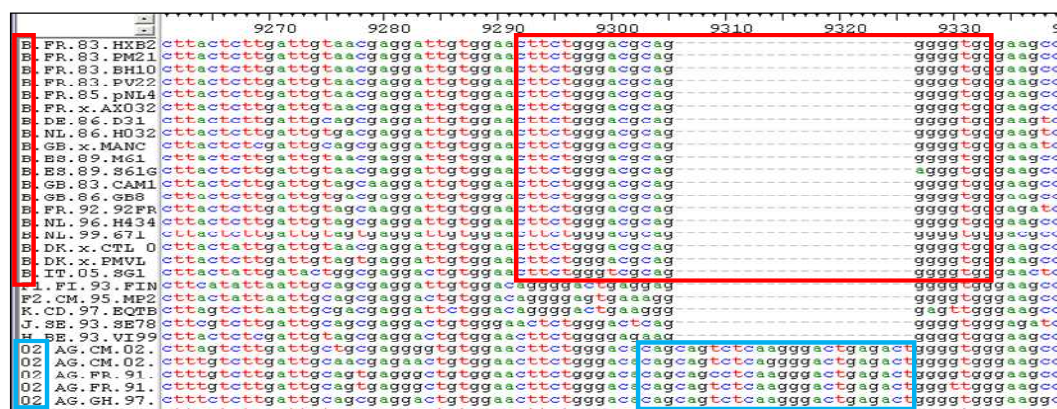
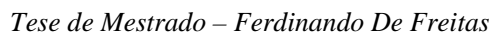


Figura 3.1 – Alinhamento múltiplo de sequências de estirpes de referência de HIV-1. A título de exemplo está ilustrada, parcialmente, uma região do genoma (região codificante da glicoproteína Gp41 do invólucro viral) onde se pode constatar as diferenças óbvias entre as sequências de vírus do subtipo B (grande rectângulo vermelho) com as sequências, por exemplo, de vírus classificados como CRF02_AG (assinalados pelo grande rectângulo azul).

A flanquear estas regiões reconheceram-se porções do genoma viral com elevado nível de conservação quando as sequências nucleotídicas dos vírus de todos os subtipos e formas recombinantes analisadas foram comparadas entre si. Estas mesmas regiões foram, então, utilizadas como alvos (sequências complementares) dos *primers* ditos universais, os quais vieram a permitir a amplificação indiscriminada (independente do subtipo/forma recombinante) dos vários subtipos e CRF. De acordo com a abordagem clássica/convencional dos ensaios de tipo MHA, foram desenhados para cada uma destas regiões conjuntos (pares) de *primers* universais externos e internos,

assim como um *primer* específico para cada um dos subtipos dominantes na epidemia portuguesa. Tendo em conta as diferentes regiões do genoma viral escolhidas como alvo (n=7), assim como os subtipos e CRF de HIV-1 que se pretendiam identificar (B, G, CRF02_AG e CRF14_BG), para as regiões de Gag, RT, Gp120 foram desenhados 3 *primers* específicos (permitindo a distinção entre vírus dos subtipos B, G/CRF14_BG e CRF02_AG), enquanto, que, para as regiões Pr, In, Rev, Gp41 foram desenhados 2 *primers* específicos (permitindo a distinção entre sequências virais dos subtipo B e G/CRF02_AG/CRF14_BG) (Figura 3.2). Alguns dos *primers* apresentam degenerações nas posições onde se verificou maior variabilidade entre as sequências comparadas. Embora os oligonucleótidos em causa tenham sido desenhados de forma a apresentarem um conteúdo C/G a variar entre os 40 e 60% e temperaturas de desnaturação (**TM**, do inglês, *Melting Temperature*) inferiores a 65°C (tal como definido no capítulo 2, secção 2.7) tal nem sempre se verificou. De facto, alguns destes vieram a apresentar temperaturas de desnaturação elevadas (iguais ou superiores ao estipulado). Estas excepções resultaram da dificuldade em cumprir com todas as disposições previamente estabelecidas, assim como pela natureza nucleotídica (polimorfismos locais) da região em estudo. A título de exemplo, a região codificante de Gp41 apresentava, naturalmente, um elevado conteúdo de G/C, enquanto que, por outro lado, a região de Gag apresentava uma elevada similaridade da sequência nucleotídica mesmo quando sequências de vírus de subtipos distintos foram comparadas entre si, obrigando à inclusão de um maior número de nucleótidos na sequência dos *primers* delineados, de modo a que fosse possível a distinção entre as sequências virais dos vários subtipos e CRF. Ambas as situações, entre outras, contribuíram para um aumento da TM nos *primers* utilizados. A sequência, coordenadas, região alvo do genoma viral e sentido (orientação relativamente à sequência codificante) dos *primers*/sondas desenhados neste trabalho encontram-se indicados no anexo 2.



3.3 Amplificações de fragmentos de DNA por PCR usando amostras de referência

Na sua formulação clássica, os ensaios de tipo MHA integram uma primeira amplificação de sequências nucleotídicas mediante a utilização de *primers* ditos universais. Esta é seguida de uma segunda amplificação em que o produto da reacção anterior é utilizado como matriz conseguindo-se, assim, a amplificação selectiva de sequências virais usando *primers* ou sondas subtipo-específicos. Neste trabalho, para além desta, foram implementadas variantes experimentais alternativas, cujos resultados se passam a descrever nos próximos capítulos. Na sua essência, estas foram desenvolvidas de forma a permitir o sucesso dos processos de amplificação, os quais são particularmente determinantes aquando do estudo de estirpes de HIV-1 de indivíduos com uma baixa carga viral. O processo de amplificação aqui proposto baseia-se na obtenção de fragmentos de genoma viral em 3 etapas sucessivas numa combinação de estratégias de *nested* e *heminested*-PCR.

3.3.1 Verificação, por PCR, da integridade estrutural das amostras de referência utilizadas

De modo a avaliar a integridade do DNA correspondente às amostras de referência, foram amplificados vários fragmentos de DNA compreendendo sequências distribuídas ao longo do genoma viral recorrendo-se, para tal, a conjuntos de *primers* utilizados com sucesso em trabalhos anteriores (ver anexo 2). A partir das diluições efectuadas na razão de 1:100, correspondendo a uma concentrações final de DNA que variou entre os $8,4 \times 10^{-3}$ e $2,6 \times 10^{-1}$ ng/ μ l foram amplificados fragmentos para 4 regiões do genoma de HIV-1 (Gag, Rt, Env, Nef) com aproximadamente 470pb, 650pb, 500pb e 770pb, respectivamente. Os resultados da amplificação foram visualizados após electroforese em gel de agarose a 1,5%, observando-se que para todas as amostras de referência foi possível amplificar, correctamente (tamanho desejado aproximado) as diferentes regiões desejadas. Foram desta “regra” excepção as amostras 8E5LAV e 3’3306, onde os produtos amplificados foram visualizados a partir de uma diluição de 1:50 correspondendo a uma concentração final na reacção de $3,4 \times 10^{-1}$ e $1,7 \times 10^{-2}$ ng/ μ l

respectivamente (Figura 3.3). A planificação do protocolo de amplificação foi realizada de modo maximizar as hipóteses de amplificação, utilizando para o efeito uma temperatura de hibridação de 50°C.

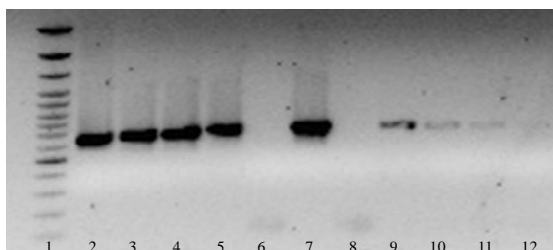


Figura 3.3 – Visualização dos produtos de PCR resultantes da amplificação da região RT (fragmento com aproximadamente ≈ 650 pb) para as amostras de referência. **1** – Marcador de massa molecular, **2** - pBD6-15 CRF02_AG(TG1), **3** - pNL4-3, **4** - p2698, **5** - 5'3037, **6** - 8E5 LAV, **7** - 5'988, **8** – Controlo negativo (água), **9** – 8E5LAV (3,4ng/ μ l), **10** – 8E5LAV (1,7ng/ μ l), **11** – 8E5LAV ($8,5 \times 10^{-1}$ ng/ μ l), **12** – 8E5LAV ($4,3 \times 10^{-1}$ ng/ μ l).

3.3.2 Amplificação de fragmentos de DNA universais externos e internos para as várias regiões em estudo

Uma vez confirmada a integridade do DNA que viríamos a utilizar como referência tal facto permitiu-nos testar a sensibilidade/especificidade dos *primers* desenhados tendo em conta as amplificações de sequências virais pretendidas (Gag, Pr, RT, In, Rev, Gp120, Gp41). Esta foi iniciada com a amplificação das diferentes regiões alvo usando *primers* universais externos e internos desenhados no âmbito deste trabalho. Usando a mesma linha de raciocínio, os fragmentos assim obtidos foram designados, respectivamente, fragmentos universais externos e internos, tal como indicado na figura 3.2. Os 3 clones contendo genomas virais parciais, assinalados com o prefixo 5', foram analisados para as regiões Gag, Pr, RT e In, ao passo que os clones restantes (designados com o prefixo 3') foram analisados para as regiões de Rev, Gp120, Gp41. As demais amostras de referência foram analisadas para todas as regiões do genoma viral em estudo (n=7).

A análise dos produtos amplificados correspondentes aos fragmentos universais externos obtidos para as várias regiões do genoma revelou, para a maioria das amostras,

um fragmento de DNA específico de tamanho aproximado ao esperado (Tabela 3.3). Para determinadas regiões, a análise em gel de agarose não revelou a presença de produto amplificado visível a partir de algumas amostras (Figura 3.4). As condições de reacção foram semelhantes às utilizadas nas amplificações preliminares (ver capítulo 2, secção 2.8.1), no entanto, foram realizadas para um volume final de 25µl, a partir de 1µl de DNA matricial onde a concentração final de DNA utilizada por reacção foi exactamente a mesma utilizada na secção anterior (ver secção 3.3.1). Para cada amostra e por cada região, as amplificações foram realizadas em triplicado correspondendo a 3 diferentes temperaturas de hibridação (55°C, 58°C e 60°C). Contudo, não se detectaram alterações no padrão de amplificação em função das três temperaturas utilizadas.

A partir dos produtos amplificados (fragmentos externos universais) acima mencionados, foram amplificados os fragmentos universais internos. Estas reacções de amplificações foram realizadas a partir de 1µl na razão de 1:1000 de DNA matricial (produto amplificado na primeira reacção, ou seja, dos fragmentos externos) ou 5µl (directamente do produto amplificado) para as amostras em que não foi possível visualizar produto amplificado na primeira reacção. A análise dos produtos amplificados revelou a presença um fragmento específico (revelando em gel de agarose um tamanho aproximado ao esperado) (Tabela 3.3) para a grande maioria das amostras e ao longo das regiões do genoma analisadas (Figura 3.4). Para as amostras pBD6-15 (clone molecular de referência para a forma recombinante CRF02_AG) não foram visualizados produtos amplificados para a maioria das regiões analisadas (fragmentos externos/internos). Para esta amostra, este facto foi obviado com base na utilização do extracto preparado a partir da estirpe bacteriana TG1 (ver capítulo 2, secção 2.2.1), Cada reacção (amostras por região) foi individualmente efectuada a 3 temperaturas de hibridação (55°C, 58°C e 60°C) tendo-se observado uma ligeira diminuição na quantidade do produto amplificado para as regiões do genoma no limite superior de temperatura utilizado (60°C). Os resultados das amplificações estão resumidos na Tabela 3.3. Em todos os conjuntos de reacções realizadas foram contemplados controlos negativos, não tendo sido alguma vez registada qualquer evidência de contaminação.

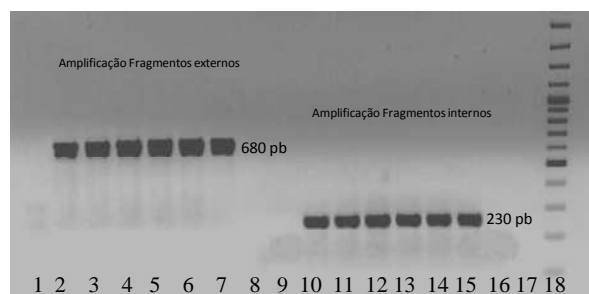


Figura 3.4 – Exemplo do perfil de visualização dos produtos amplificados para os fragmentos externos e internos, neste caso para a região da Pr (55°C). Fragmentos externos (≈ 680 pb): **1** – pBD6-15 CRF02_AG(XL1), **2** - pBD6-15 CRF02_AG(TG1), **3** – p2698, **4** – 5'3037, **5** – 5'3006, **6** – 5'988, **7** – pNL4-2, **8** - Controlo negativo (água); Fragmentos internos (≈ 230 pb): **9** – pBD6-15 CRF02_AG(XL1), **10** - pBD6-15 CRF02_AG(TG1), **11** – p2698, **12** – 5'3037, **13** – 5'3006, **14** – 5'988, **15** – pNL4-2, **16** - Controlo negativo **17** – pista livre **18** – Marcador de massa molecular.

Tabela 3.3 – Resumo dos resultados obtidos para a amplificação dos fragmentos externos e internos para as amostras de referências por todas as regiões analisadas do genoma. Estão ainda representados o tamanho relativo dos fragmentos (externo e interno) esperados para cada região, bem como, a concentração à qual as amostras foram amplificadas. N - regiões para as quais não existiam correspondência no DNA alvo.

	Região alvo	Gag		Protease		RT		Integrase		Rev		Gp120		Gp41	
		Ex	In	Ex	In	Ex	In	Ex	In	Ex	In	Ex	In	Ex	In
Amostra de Referência	Tamanho \approx (pb)	620	230	680	230	670	280	600	210 275	1025	270	1025	140 510	850	200
Diluição															
CRF02_AG (XL1)	1:100	-	+	-	-	-	-	-	-	-	-	-	-	-	-
CRF02_AG (TG1)	1:100	+	+	+	+	+	+	-	+	-	+	-	+	-	+
p2655	1:100	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pNL4-3	1:100	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5'3037	1:100	+	+	+	+	+	+	-	+	N	N	N	N	N	N
3'3037	1:100	N	N	N	N	N	N	N	N	+	+	+	+	+	+
5'3306	1:100	+	+	+	+	+	+	-	+	N	N	N	N	N	N
3'3306	1:5	N	N	N	N	N	N	N	N	-	+	-	+	-	+
5'588	1:100	+	+	+	+	+	+	-	+	N	N	N	N	N	N
3'588	1:100	N	N	N	N	N	N	N	N	+	+	+	+	+	+
8E5LAV	1:5	-	+	+	+	-	+	+	+	-	+	-	+	-	+

3.3.3 Amplificação específicas (subtipo/CRF) de diferentes regiões alvo distribuídas ao longo do genoma do HIV-1

Na secção anterior foram apresentados dados que nos permitiram constatar a possibilidade de amplificação das diferentes regiões alvo do genoma do HIV-1 mediante a utilização de *primers* universais externos e internos como iniciadores de PCR. Esta foi traduzida pela visualização, em gel de agarose, de fragmentos de amplificação (amplicões) específicos. Na consequência dos resultados obtidos foram, então, testadas condições em que a utilização combinada de um *primer* específico e um outro universal (ver Figura 2.1, Capítulo 2 secção 2.8.2 e secção 3.3) permitissem a amplificação selectiva, dependente agora da sua identidade genética (subtipo/CRF), de cada uma das regiões do genoma viral em análise. Este passo baseou-se na optimização prévia de condições experimentais, mais uma vez baseada na utilização de um conjunto de amostras de referência. Todas as combinações (amostra/região) foram efectuadas num formato de 6 reacções individuais, nas quais se variou a temperatura de hibridação (55, 58, 60, 63, 66 e 69°C). Em todas as reacções foram contemplados a preparação de reacções de controlo negativo tendo estas, servido, essencialmente, para controlar a possível ocorrência de contaminações.

Para a maioria das regiões (Gag, Pr, RT, Rev e Gp41), e para todas as amostras de referência utilizadas, foi possível constatar a amplificação específica de um fragmento de dimensões aproximadas ao esperado, a pelo menos, uma das temperaturas de hibridação utilizadas. Por um lado, para a maioria das regiões observou-se uma redução do produto amplificado quando as reacções decorreram à temperatura de hibridação de 69°C; por outro lado, para alguns ensaios verificaram-se amplificações inespecíficas às temperaturas mais baixas. Como exemplo disto, para a região codificante de Gp41 o *primer* específico para o subtipo G às temperaturas de hibridação de 55 e 58°C permitiu, ainda que de uma forma pouco evidente (produto obtido em quantidade menor do que o amplicão específico foi obtido a partir da mesma matriz de DNA), a amplificação de sequências virais a partir das amostras de referência do subtipo B. Por esta razão, a temperatura escolhida para a realização do ensaio para esta região foi intermédia, tendo sido seleccionada, neste caso, 63°C (Figura 3.5). As temperaturas para as quais se obtiveram resultados de amplificação específicos e o

tamanho esperado dos fragmentos para cada região analisada estão sumarizados na Tabela 3.4.

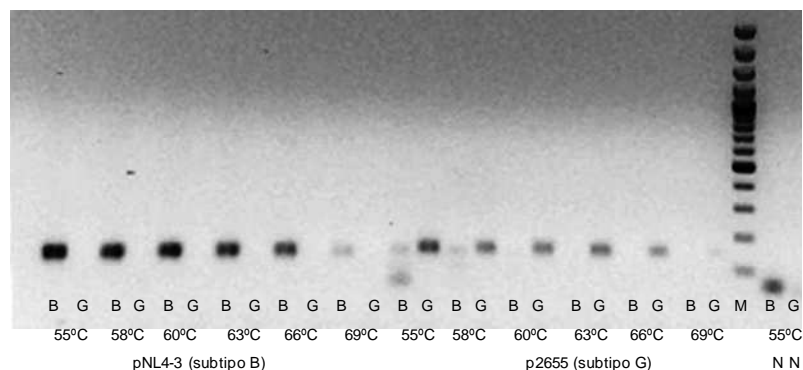


Figura 3.5 – Exemplo do perfil de amplificação específica obtido para duas amostras de referência (pNL4-3 -subtipo B e p2655-subtipo G). Os fragmentos amplificados referem-se a parte da região que codifica a glicoproteína Gp41. São ainda reconhecidas diminuições na quantidade de produto amplificado como função das temperaturas de hibridação utilizadas. N - Controle negativo; M – marcador de massa molecular.

Para a região do genoma viral que codifica a integrase, a análise dos produtos para as 8 amostras de referência utilizadas revelou a amplificação de um fragmento de DNA de pequenas dimensões (resolvido apenas em gel de poliacrilamida) independentemente da temperatura de hibridação analisada. Perante esta situação, optou-se ainda por delinear um novo *primer* interno de modo a que o fragmento desejado apresentasse maiores dimensões. No entanto, os produtos obtidos revelaram um padrão de amplificação inespecífico. Tal foi traduzido pelo facto de que o *primer* específico para sequências virais do subtipo B permitiu a hibridação heteróloga (com sequências de referência do subtipo G e da forma recombinante CRF02_AG) à temperatura máxima de hibridação utilizada (69°C). Face a esta situação, optou-se por redesenhar um novo conjunto de *primers* específicos, tendo como alvo uma região próxima a montante da anterior. No delineamento dos *primers* específicos foi contemplada, de forma propositada, a inclusão de nucleotídeos não homólogos na sequência do *primer* específico para as sequências do subtipo G e forma recombinante CRF02_AG. A análise dos produtos amplificados obtidos mediante a utilização deste segundo grupo de *primers* revelou que o *primer* específico para o subtipo B permitia a

amplificação de fragmentos heterólogos (subtipo G e CRF02_AG) e que o *primer* específico para as sequências do subtipo G originava fragmentos de tamanhos superiores aos esperados quando a matriz de amplificação utilizada correspondia a DNA do clone molecular correspondendo à forma CRF02_AG. Estes resultados revelaram-se constantes apesar da grande amplitude de temperaturas de hibridação utilizadas (de 55 a 69°C).

Relativamente à região que codifica parte da glicoproteína Gp120, a análise em gel de agarose dos produtos de amplificação obtidos revelou que, independentemente da amostra de referência e temperatura de hibridação utilizada, os *primers* de amplificação específica (GP120-14B/G/CRF02_AG sufixo 1) originavam, de uma forma geral, produtos de amplificação independentemente da identidade genética da matriz de referência utilizada, isto é, estes tanto eram amplificados a partir de amostra de referência homóloga (ex: *primer* B/referência de subtipo B) como heteróloga (ex: *primer* G/referência de subtipo B) relativamente ao *primer* específico utilizado. Contudo, as pequenas dimensões (<50pb) dos fragmentos amplificados dificultavam a correcta visualização de possíveis fragmentos específicos. Face a esta situação foram redesenhados 6 novos *primers* específicos (dois conjuntos: subtipo Gp120-14B/G/CRF02_AG – sufixo 2 e 3) juntamente com dois novos *primers* de amplificação universal. Os novos *primers* de amplificação específica foram delineados de forma a hibridar com regiões alvo distintas da definida anteriormente, tendo sido orientados no sentido reverso relativamente aos primeiros (e à sequência codificante). Estes conjuntos diferiam, essencialmente, no número de nucleótidos, onde a sequência dos *primers* Gp120-14B/G/CRF02_AG 2 compreendia a sequência dos *primers* Gp120-14B/G/CRF02_AG 3. Contudo, neste último conjunto foi propositadamente contemplada a inclusão de nucleótidos não homólogos (i.e. *a priori* não complementares à sequência alvo) na sequência dos *primers* de modo a desfavorecer hibridações com as sequências virais de vírus de um determinado subtipo, relativamente a outros.

Após realização, com sucesso, das amplificações (a 55°C) dos fragmentos universais internos (recorrendo ao novo conjunto de *primers* universais internos acima mencionados) procedeu-se à realização das reacções de amplificação de carácter específico. A análise dos fragmentos amplificados com o segundo conjunto Gp120-

14B/G/CRF02_AG 2 revelou a presença de amplicões específicos quando o *primer* específico para sequências do subtipo B foi utilizado em conjunto com uma matriz do mesmo subtipo. Por seu lado, o *primer* específico para sequências do subtipo G originou amplicões quando utilizadas como matriz amostras de referência do subtipo G, ou da forma CRF02_AG. No entanto, para as reacções de amplificação em que foram utilizadas temperaturas de hibridação superiores a 60°C, verificou-se uma diminuição visível (em gel de agarose) da quantidade do produto amplificado quando a matriz utilizada e *primer* específico era não-homólogos (ex: *primer* G/referência de subtipo B ou CRF02_AG). De uma forma semelhante, o *primer* específico para sequências da forma recombinante CRF02_AG permitiu a obtenção às temperaturas de 55 e 60°C de amplicões, independentemente das referências utilizadas corresponderem a sequências dos subtipos B ou G. Para temperaturas de hibridação iguais ou superiores a 63°C não se visualizavam produtos de amplificação para as referências do subtipo B, contrariamente ao resultado observado quando utilizada a referência de subtipo G tendo-se verificado amplificações heterólogas até à temperatura máxima experimentada (69°C). Para o terceiro conjunto de *primers* específicos (Gp12014B/G/CRF02_AG 3) a análise dos produtos de amplificação revelou a presença de um amplicão homólogo revelado para cada um dos *primers* específicos (subtipo B e CRF02_AG) e respectivas amostras (à temperatura de hibridação de 63°C). Contudo, simultaneamente, foi detectada a presença de fragmentos inespecífico de dimensões superiores ao esperado em todas as reacções de amplificação, incluindo as que foram realizadas à temperatura máxima utilizada ao longo destes ensaios (69°C). Para *primer* específico para o subtipo G foram detectados amplicões quando utilizados com as amostras do subtipo G e CRF02_AG. As temperaturas para as quais se obtiveram resultados de amplificação específicos e o tamanho esperado dos fragmentos para cada região analisada estão representadas na Tabela 3.4.

Tabela 3.4 – Temperaturas e tamanho do fragmento esperado de acordo com o *primers* subtipo/CRF específicos utilizados e respectiva região do genoma. Estão ainda representados as várias combinações utilizadas para região da Integrase e Gp120. Para a região da Integrase a notação 290* representa o tamanho do fragmento esperado a quando a utilização de um novo *primer* universal interno. N-não aplicável.

Região alvo	Primer específico	Tamanho (≈pb)	Temperatura (°C)
Gag	B/_02/G	150	69
Pr	B/G	150	58
Rt	B/_02/G	130	60
In	B-1	40/290*	N
	G-1	40/290*	N
	B-2	250	N
	G-2	250	N
Rev	B/G	170	58
Gp120	B-1/_02-1/G-1	30	N
	B-2	400	63
	_02-2	350	63
	G-2	450	63
	B-3	400	63
	_02-3	360	63
	G-3	460	63
Gp41	B	108	63
	G	95	63

3.3.4 Amplificação de sequências genómicas de HIV-1 por PCR em tempo-real: ensaio de competição de matrizes

A especificidade dos *primers* internos foi ainda testada através de um ensaio de competição de matrizes. Basicamente, nestes procedeu-se à amplificação de diferentes regiões do genoma viral em estudo utilizando amostras de referência e uma abordagem experimental baseada na PCR em tempo-real. Uma vez que as amplificações específicas descritas na secção anterior apresentavam resultados semelhantes para todas as amostras de um mesmo subtipo/CRF, optámos pela escolha de um “representante” de cada subtipo/CRF testado (subtipo B-pNL4-3, subtipo G-p2655, CRF02_AG-pBD6-15). As

temperaturas de hibridação utilizadas foram aquelas para as quais, e tal como descrito na secção anterior, foram obtidos os melhores resultados de amplificação. Os ensaios foram realizados a partir de fragmentos universais externos (ver secção 3.3.2). Contudo, antes de se terem iniciado as amplificações propriamente ditas, os produtos amplificados foram quantificados por espectrofotometria de modo a que as diferentes matrizes utilizadas se apresentassem, nas diferentes reacções de amplificação efectuadas, numa concentração aproximadamente semelhante por cada região analisada (ver secção 3.1). Os ensaios foram elaborados de acordo com o formato definido (Capítulo 2, secção 2.8.4). Resumidamente, pretendeu-se avaliar a eficiência de ligação de cada um *primers* específicos às respectivas sequências alvo homólogas, e a forma como esta poderia ser influenciada pela presença de DNA competidor em concentração idêntica ou superior à da matriz homóloga, correspondendo à mesma região do genoma viral, mas de natureza heteróloga relativamente ao *primer* específico utilizado. Assim, e em simultâneo, foram preparadas 3 reacções de PCR individuais (DNA competidor heterólogo/DNA alvo homólogo), nas quais a concentração do DNA competidor foi mantida constante, ao passo que concentração da matriz homóloga variou de forma a que, relativamente ao DNA competidor, este último se apresentasse nas razões (DNA competidor:DNA homólogo) 1:1, 10:1 e 100:1. Estes ensaios foram realizados na presença de *SYBR[®] Green I* na mistura reaccional. Para cada conjunto de reacções contemplou-se, ainda, a realização de um controlo positivo de reacção (presença exclusiva de DNA alvo homólogo relativamente ao *primer* específico utilizado) e um controlo negativo (destinado ao registo de eventuais contaminações).

A análise das curvas de amplificação demonstrou que para todas as regiões (Gag, Pr, RT, Rev, Gp120 e Gp41) (Figura 3.6), apesar do competidor (DNA não homólogo) estar presente em concentrações semelhantes ou superior relativamente à matriz homóloga, os *primers* de amplificação específica permitiram a correcta detecção da mesma.

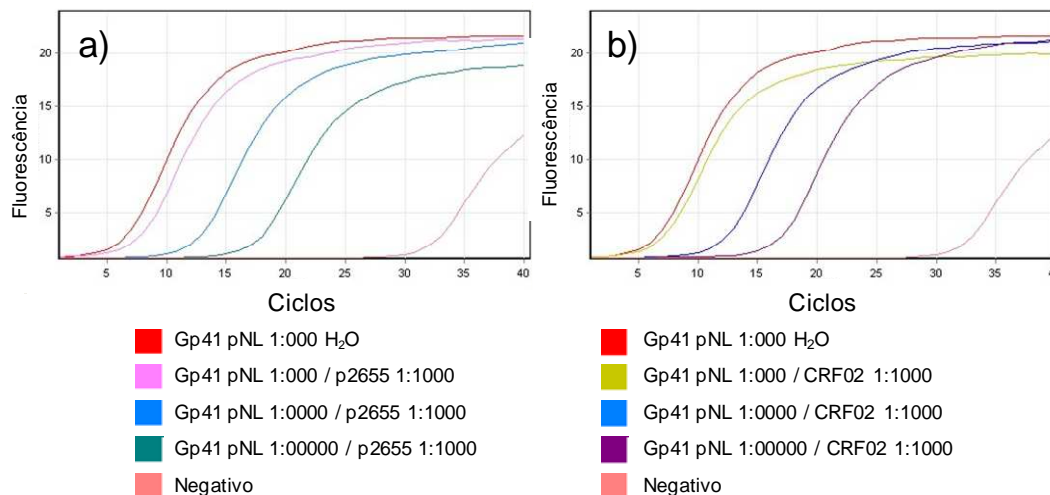


Figura 3.6 – Exemplo de um ensaio de competição entre duas matrizes de referência (região de Gp41). a) O conjunto de *primers* específico para o subtipo B permitiu a amplificação específica da amostra pNL4-3 (subtipo B) na presença do competidor p2655 (subtipo G); b) Amplificação específica da mesma amostra pelo mesmo par de *primers* mas na presença do competidor CRF02_AG (subtipo G). As curvas a vermelho representam a amplificação da amostra sem competidor (controlo positivo), sendo que as restantes resultam da amplificação do DNA “alvo” ao longo das várias diluições e controlo negativo (rosa claro).

3.4 Amplificação por PCR de sequências virais alvo a partir de amostras clínicas

3.4.1 Obtenção de fragmentos universais externos e internos

Uma vez preparados os extractos de RNA total obtidos a partir de alíquotas de plasma das amostras clínicas analisadas neste trabalho utilizaram-se, na maioria dos casos, 5µl das reacções de transcrição reversa (contendo moléculas de cDNA) para a amplificação de sequências virais sob a forma de fragmentos universais. Em alternativa, as amplificações foram efectuadas utilizando-se dois pequenos discos contendo DNA numa matriz sólida, correspondendo estas situações às amostras clínicas disponíveis sob a forma de sangue total imobilizado em cartões FTA™. A amplificação por *nested*-PCR dos fragmentos universais externos e internos, para todas as regiões analisadas, foi realizada à temperatura de hibridação de 55°C. A análise em gel de agarose dos produtos de amplificação obtidos revelou, em particular, para as regiões de Gag e Pr a presença dos fragmentos de tamanho aproximado ao esperado (Tabela 3.4 secção 3.3.2)

no primeiro passo de amplificação (fragmentos externos) e para a grande maioria das amostras biológicas. Contudo, para algumas amostras, não foram detectados produtos amplificados visíveis, facto que se verificou, como regra geral, para as restantes regiões analisadas (RT, Rev, Gp120 e Gp41) (Figura 3.7 a) e que justificou a abordagem de *nested*-PCR inicialmente planeada. A região alvo que incluiu parte da sequência codificante da Integrase não foi analisada, pelas razões já descritas anteriormente. As reacções de amplificação dos fragmentos universais internos foram iniciadas a partir de 1µl de uma diluição na razão de 1:1000 do produto amplificado (fragmento universal externo) quando este foi detectado nas primeiras reacções de amplificação. Em alternativa estas foram iniciadas utilizando 5µl das mesmas no caso das amostras para as quais a análise de alíquotas das primeiras amplificações não revelou, em gel de agarose, qualquer amplificação específica. A análise dos produtos amplificados referentes aos fragmentos universais internos relativos às várias regiões do genoma em análise (salvo a região In) revelou que, para a maioria das amostras, foi possível amplificar um fragmento de tamanho aproximado ao esperado (Figura 3.7 b).

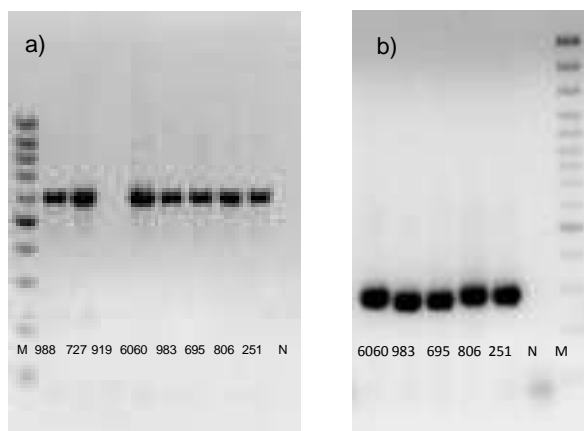


Figura 3.7 – Observação em gel de agarose dos produtos de amplificação para as amostras clínicas analisadas, referidos no texto como fragmentos universais externos e internos. a) Identificação de produtos de amplificação relativos à região RT (fragmentos externos); a análise de uma alíquota da reacção referente à amostra 919 não revelou qualquer amplificação. b) Exemplo de um perfil de amplificação para a região Rev (fragmentos internos). N - controlo negativo; M – marcador de massa molecular.

3.4.2 Amplificação por PCR de sequências virais específicas

A planificação do protocolo experimental destinado à amplificação dos fragmentos específicos a partir de amostras clínicas foi realizada tendo em conta as temperaturas de hibridação optimizadas com base na utilização de amostras de referência (Tabela 3.4, secção 3.3.3). Esta etapa surgiu na sequência da análise dos resultados apresentados na secção anterior. Assim, após terem sido amplificados os fragmentos universais internos para todas as regiões do genoma das diferentes estirpes de HIV-1 presentes nas amostras clínicas em estudo, estes foram diluídos na razão de 1:1000 (relativamente à sua concentração original nas reacções do segundo passo de amplificação contendo fragmentos universais internos gerados por *nested*-PCR), sendo que 1µl foi usado como matriz para as reacções de amplificação específica.

Tal como até agora descrito, alíquotas de todas as reacções de amplificação específicas foram analisadas por electroforese em gel de agarose de forma a verificar a amplificação específica desejada e/ou a ocorrência de amplificações inespecíficas. Os resultados obtidos revelaram, de uma forma geral, que os *primers* específicos originavam amplificações indiscriminadas para todas as regiões excepto para a região de Gag. Para esta região foi possível obter um perfil de amplificação discriminatório para 39 das 50 amostras (78%), 5 amostras revelaram um perfil de amplificação duplo (amplificação heteróloga a partir de dois dos *primers* específicos utilizados) e as restantes apresentavam amplificações indiscriminadas como consequência da utilização dos 3 *primers* específicos utilizados. Para as restantes regiões analisadas, e a partir de um conjunto de amostras biológicas seleccionadas aleatoriamente de entre o total em estudo, procedeu-se a uma nova optimização das reacções de amplificação específica. Para tal foram efectuados incrementos sucessivos (1°C) na temperatura de hibridação referente a cada um dos protocolos de PCR de modo a que o resultado permitisse uma situação de compromisso entre o número máximo de amplificações específicas e o número mínimo de resultados nulos, isto é, amplificações limitadas pelas elevadas temperaturas de hibridação utilizadas. Face aos resultados obtidos foi possível definir um novo conjunto de temperaturas de hibridação para cada região analisada. As temperaturas estão representadas na Tabela 3.5.

Tabela 3.5 – Temperaturas de hibridação utilizadas na amplificação por PCR de sequências virais específicas, re-otimizadas com base nos resultados de amplificação obtidos para amostras clínicas. Estão ainda representados os tamanhos dos fragmentos esperados de acordo com o *primers* subtipo/CRF específicos utilizados e respectiva região do genoma.

Região alvo	Primer específico	Tamanho (≈pb)	Temperatura (°C)
Gag	B/_02/G	150	69
Pr	B/G	150	63
Rt	B/_02/G	130	61
Rev	B/G	170	63
Gp120	B-2/B-3	400/400	66
	_02-2/_02-3	350/360	66
	G-2/G-3	450/460	66
Gp41	B	100	68
	G	100	68

A análise dos produtos amplificados referentes às regiões Pr, RT, Rev, Gp120 e Gp41 revelou para a maioria das amostras, perfis de amplificação específicos exclusivamente para um dos *primers* de amplificação específica por cada região (*primer* específico para o subtipo B, G ou CRF02_AG quando aplicável, ver 3.3, secção 3.2). Para a região Pr, 36 das 50 amostras (72%) apresentaram um resultado de amplificação exclusiva para um dos *primers* (subtipo B ou G), sendo que as restantes apresentaram um perfil duplo de amplificação, apesar de se terem verificado, em algumas situações, variações significativas na quantidade de produto amplificado. A Figura 3.8 representa um perfil de amplificação característico dos ensaios realizados para esta região.

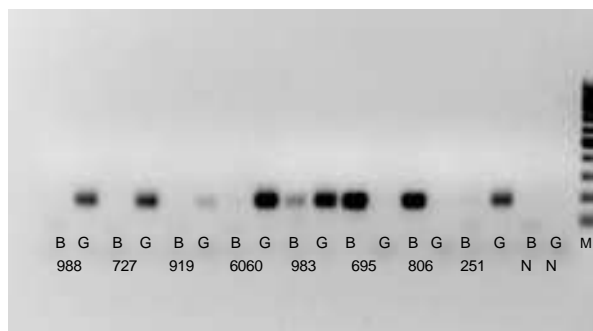


Figura 3.8 – Amplificação dos fragmentos específicos referentes a um conjunto de amostras clínicas para a região codificante de parte da protease viral. Estes ensaios foram realizados à temperatura de hibridação de 63°C, neste conjunto em particular apenas uma amostra (983) apresenta um perfil duplo de amplificação. N - Controlo negativo; M - marcador de massa molecular.

Para a região RT os resultados das amplificações específicas revelaram que um conjunto de 40 amostras (80%) foi possível obter fragmentos visíveis apenas para um dos *primers* específicos utilizados, 7 amostras com resultado duplo e para as restantes 3 amostras não foi detectado qualquer produto amplificado independente do *primer* específico. Para a região Rev, 40 amostras (80%) apresentaram um perfil de amplificação discriminatório, 6 amostras apresentaram um perfil de amplificação duplo, e para as restantes amostras não foi observado em gel de agarose qualquer produto de amplificação para qualquer um dos *primers* específicos. Para a região Gp120, a análise dos produtos resultantes de amplificações específica efectuadas com base na utilização do segundo conjunto de *primers* desenhados revelou, para 34 das amostras clínicas em estudo (68%), um perfil de amplificação exclusivo com um único *primer* de amplificação específica, 7 apresentavam um perfil de amplificação duplo, 3 apresentavam um perfil de amplificação indiscriminado originado pela amplificação a partir de todos os *primers* específicos (resultado triplo), enquanto que, 5 apresentavam um perfil inespecífico, traduzido pela presença em simultâneo de várias amplicões de dimensões não esperadas. Finalmente, para uma única amostra, não foram detectados quaisquer produtos de amplificação. Tendo em conta estes resultados, optámos por reanalisar esta mesma região do genoma viral, testando a capacidade que o terceiro conjunto de *primers* específicos teria em dar origem aos amplicões desejados. A análise efectuada demonstrou que das 50 amostras analisadas, 36 permitiram a verificação de amplificações correspondentes ao fragmento de tamanho aproximadamente ao esperado

exclusivas com um dos *primers*, 7 apresentavam amplificação dupla ou tripla, 5 apresentavam produtos amplificados inespecíficos (fragmentos de dimensões não esperadas), enquanto que para duas amostras não foi detectado qualquer produto de amplificação. No entanto, para todas as combinações foram ainda detectadas amplificações inespecíficas, reveladas pela presença de fragmentos de DNA de dimensões diferentes às esperadas para esta região (Tabela 3.4). Para a região Gp41 foram identificadas 39 amostras (78%) caracterizadas pela presença de um fragmento específico e exclusivo para apenas um dos *primers* utilizados (específico para o subtipo B ou G), duas amostras apresentaram um resultado duplo e para as restantes não foi detectado produto amplificado. Os resultados das amplificações específicas para cada região e amostra estão resumidos na Tabela 3.6.

Tabela 3.6 – Resumo dos resultados obtidos para as amplificações específicas a partir das amostras clínicas ao longo das regiões analisadas. Os dados estão representados pela frequência de amplificação específica (apenas para um dos *primers* subtipo/CRF específico) e respectiva percentagem sobre o total de amostras analisadas. Estão também representados para cada região a frequência de reacções inespecíficas (amplificação por dois ou três *primers* subtipo/CRF específicos) e reacções para as quais não foram detectados produtos amplificados.

Amplificações					
Região	Específicas	(%)	Inespecíficas	Não detectadas	Total
Gag	39	78	11	0	50
Pr	36	72	14	0	50
Rt	40	86	7	3	50
Rev	40	80	6	4	50
Gp120-2	34	68	15	1	50
Gp120-3	36	72	12	2	50
Gp41	39	78	2	9	50

3.4.3 Utilização do PCR em tempo-real para amplificação específica de sequências de HIV-1 e detecção, em simultâneo, dos produtos de amplificação

Na sua aborgadem convencional, os métodos de genotipagem de HIV-1 baseados em ensaios do tipo MHA fazem uso, na sua generalidade, de sondas de hidrólise subtipo-específicas, as quais são utilizadas para a detecção exclusiva de produtos de amplificação num formato de PCR em tempo-real. Neste trabalho optou-se, inicialmente, por uma abordagem mais simples e menos dispendiosa, na qual a detecção em tempo-real dos fragmentos específicos amplificados com base na utilização de *primers* específicos para cada um dos subtipos/CRF, foi feita na presença de *SYBR*[®] *Green I*.

No seguimento das amplificações específicas por PCR convencional (e análise de resultados em gel de agarose) e recorrendo à informação da re-optimização (novas temperaturas de hibridação) das condições específicas de reacção (Tabela 3.5 secção 3.4.2) realizaram-se os ensaios de amplificação numa plataforma de PCR em tempo-real. As amplificações foram registadas como específicas quando para cada amostra analisada foi apenas observada uma curva referente a uma amplificação exponencial em pelo menos 5 ciclos consecutivos a partir do conjunto de *primers* específicos disponíveis para cada região, sem que fossem registadas amplificações exponenciais a partir dos restantes *primers* nos 10 ciclos seguintes.

As amplificações foram conduzidas a partir de uma diluição de 1:1000 do produto obtido na segunda reacção de amplificação, correspondente à amplificação dos fragmentos universais internos. Em todas os ensaios foram contemplados controlos negativos, para os quais, não se registaram sinais fluorescentes compatíveis com amplificações provenientes de possíveis contaminações acidentais durante o processo de preparação/manipulação das diferentes amostras. De igual forma, a análise das curvas referentes ao controlo interno de reacção evidenciaram que os resultados obtidos não foram, à partida, influenciados por erros de pipetagem.

A análise das curvas de amplificação revelou para a maioria das amostras e para as várias regiões analisadas, uma amplificação específica restrita à utilização de um dos *primers* específicos (76 a 90% do total de amostras, dependendo da região estudada). No entanto, para algumas amostras foi registada uma amplificação indiscriminada por

mais do que um *primer* e para um grupo reduzido de amostras e regiões não se registou amplificação com qualquer um dos *primers* específicos utilizados. (ver anexo1). A região Gp120 foi aquela para a qual foi registado o menor número de amplificações de fragmentos específicos, apenas 38, sendo que as restantes foram amplificadas indiscriminadamente por mais do que um *primer* específico. Por outro lado, por entre todas as regiões estudadas, a região Pr foi a região que apresentou um maior número de amplificações específicas, reveladas pelo registo de 45 fragmentos específicos (uma única curva de amplificação exclusiva tendo em conta os *primers* específicos utilizados) num total de 50 amostras em estudo. Na Figura 3.9 estão representados, a título de exemplo, dois ensaios de amplificação realizados no decorrer deste trabalho para as regiões Pr e Gp120. A Tabela 3.7 resume, para cada uma das regiões do genoma viral analisadas, o conjunto de resultados obtidos.

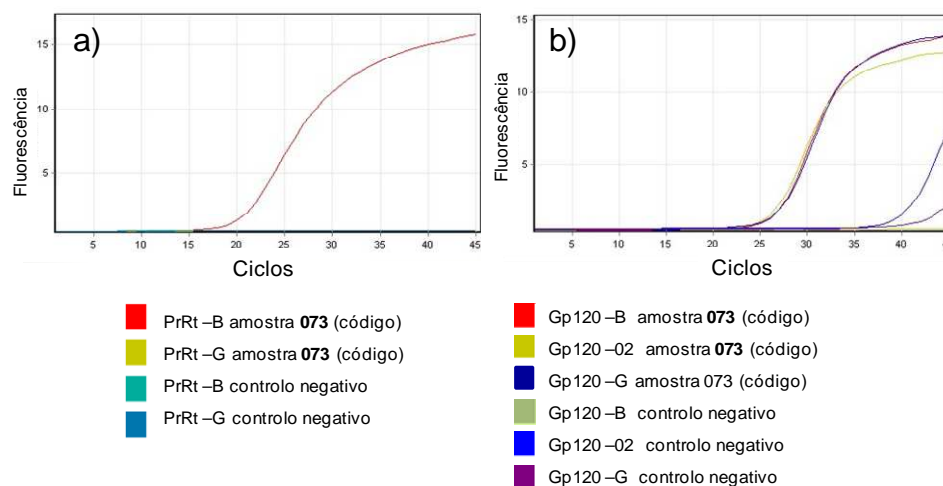


Figura 3.9 – Ensaio de amplificação específica por PCR em tempo-real. a) Região Pr: a curva de exponencial (vermelho) demonstra a amplificação discriminatória neste caso da amostra “073” pelo *primer* específico para o subtipo B. Por outro lado a curva exponencial referente à utilização do *primer* específico para sequências virais do subtipo G (amarelo mostarda) apresenta uma fluorescência equivalente às curvas apresentadas para os controlos negativo para ambos os subtipos que poderá ter origem na formação de dímeros ou outras tipo de moléculas de DNA em cadeia dupla resultado de auto-hibridação e amplificação de *primers* entre si. b) Região de Gp120: As curvas resultam de uma amplificação inespecífica para uma amostra biológica a partir dos *primers* subtipo/CRF específico.

Tabela 3.7 – Resumo dos resultados obtidos para as amplificações específicas por PCR em tempo-real a partir das amostras clínicas ao longo das regiões analisadas. Os dados estão representados pela frequência de amplificação específica (apenas para um dos *primers* subtipo/CRF específico) e respectiva percentagem sobre o total de amostras analisadas. Estão também representados para cada região a frequência de reacções inespecíficas (amplificação por dois ou três *primers* subtipo/CRF específicos) e reacções para as quais não foram detectados produtos amplificados.

Amplificações (PCR em tempo-real)					
Região	Específicas	(%)	Inespecíficas	Não detectadas	Total
Gag	42	84	8	0	50
Protease	45	90	5	0	50
RT	44	88	5	1	50
Rev	44	88	2	4	50
Gp120-2	38	76	12	0	50
Gp41	44	88	0	6	50

Alternativamente a estes ensaios, que como se descreveu anteriormente se basearam na utilização, como matriz, dos produtos das segundas reacções de amplificação universais, foi ainda avaliada a possibilidade de utilização do produto amplificado na primeira reacção (fragmentos universais externos) como matriz para os ensaios de amplificação específica. Estes ensaios foram executados, para as várias regiões analisadas, não a partir da totalidade mas de um grupo de amostras escolhidas aleatoriamente de entre o conjunto total em estudo. Para as reacções de amplificação foram utilizados, como DNA matricial, 1µl de diluições 1:1000 para as amostras que apresentaram produtos de amplificação após o primeiro ciclo de PCR ou, em alternativa 5µl para as que não apresentaram fragmentos observáveis em gel de agarose. A análise dos resultados demonstrou a amplificação de fragmentos específicos para as regiões de Gag, Pr, e RT. Contudo, para as restantes regiões (Rev, Gp120 e Gp41) observaram-se inesperadamente curvas de amplificações inespecíficas a partir dos *primers* subtipo/CRF específicos para todos os ensaios realizados (os resultados não foram confirmados através da análise em gel de agarose de alíquotas da reacções de amplificação) (Figura 3.10).

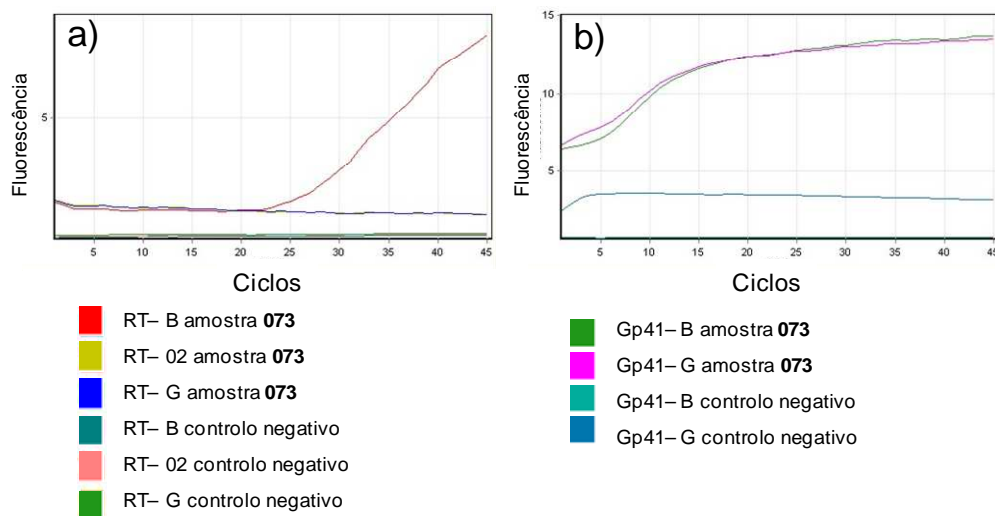


Figura 3.10 – Ensaio de amplificação específica por PCR em tempo-real. a) Região RT: A curva de amplificação exponencial (vermelho) demonstra a amplificação discriminatória neste caso da amostra “073” pelo *primer* específico para o subtipo B. b) Região Gp41: As curvas resultam de uma amplificação inespecífica para uma amostra biológica “073” a partir dos *primers* subtipo/CRF específico B e G.

3.4.4 Detecção de amplificação específica de sequências virais (Gp41) de HIV-1 por PCR em tempo-real usando sondas de hidrólise (*TaqMan*)

As amplificações específicas de sequências de HIV-1 presentes nas amostras clínicas analisadas foram ainda efectuadas para a região do genoma que codifica parcialmente a glicoproteína Gp41 tendo-se recorrido, para tal, à PCR em tempo-real mas, desta feita, socorrendo-nos de sondas de hidrólise para a detecção dos produtos de amplificação pretendidos. As condições de reacção foram semelhantes às utilizadas na secção anterior para a detecção de amplicões cobrindo a mesma região do genoma viral, nomeadamente no que diz respeito à temperatura de hibridação utilizada nos ciclos de PCR (68°C), uma vez que a esta permitiu a amplificação discriminatória de sequências virais das amostras analisadas. O protocolo experimental utilizado diferiu do anterior, essencialmente, pela utilização simultânea de dois canais de leitura de fluorescência distintos, uma vez que as sondas específicas para estirpes virais dos subtipos B e G

foram propositadamente marcadas com fluoróforos diferentes: JOE e ROX, respectivamente. As sondas específicas foram desenhadas de forma a apresentarem, exactamente, a mesma sequência nucleotídica que os *primers* específicos, utilizados na secção anterior para a amplificação selectiva desta mesma região do genoma viral. Os ensaios foram efectuados a partir de 5µl das reacções de amplificação contendo fragmentos universais externos (secção 3.4.1) recorrendo-nos, para tal, do uso dos *primers* universais internos e de cada uma das sondas específicas. A análise das curvas de amplificação para 27 amostras, escolhidas aleatoriamente de entre as analisadas neste trabalho, revelou que para 14 (52%) destas foi possível detectar um sinal de ligação específico para um dos subtipos virais (B ou G) (Figura 3.11). Para as restantes amostras, inesperadamente, não foram registados quaisquer sinais indicadores de amplificação de sequências virais, independentemente da sonda utilizada. Paralelamente, foram ainda efectuados ensaios de tipo *multiplex* no qual as duas sondas específicas foram utilizadas simultaneamente na mesma reacção de amplificação. A detecção de sinal foi efectuada da mesma forma que anteriormente (simultaneamente em canais diferentes) e foram escolhidas duas das amostras para as quais foram obtidos resultados específicos distintos referentes à amplificação de sequências virais para o subtipo B e para o subtipo G. Os resultados obtidos neste tipo de ensaio demonstraram a detecção de sinal específico para cada uma das amostras em análise tendo, para ambas, sido registado o mesmo padrão revelado nos ensaios realizados anteriormente. É importante referir que este ensaio foi realizado, a título experimental, para um número reduzido de amostras. Os resultados obtidos encontram-se resumidos juntamente com os resultados obtidos a partir das várias estratégias utilizadas neste trabalho no anexo 1.

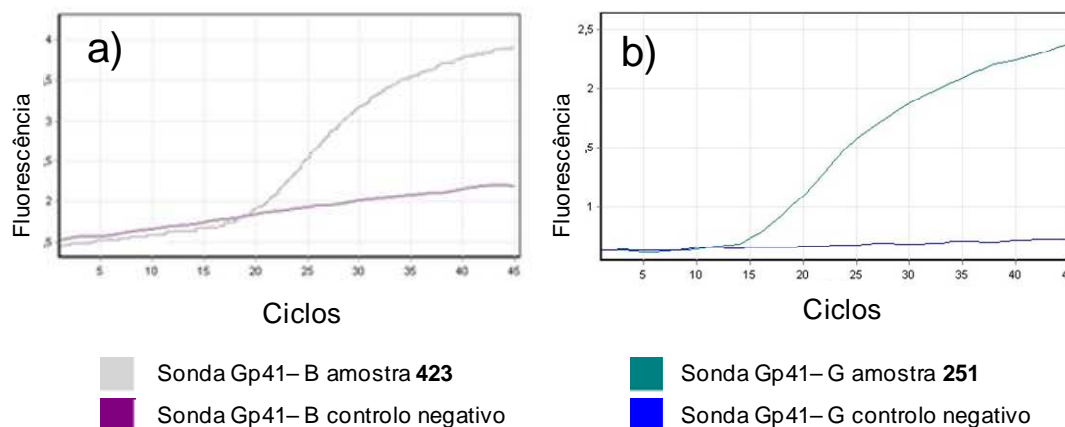


Figura 3.11 – Ensaio de amplificação específica por PCR em tempo-real com sonda de hidrólise (TaqMan) para a região Gp41 em canais de detecção independentes. a) Amplificação específica de uma amostra (423) através da sonda específica para o subtipo B (JOE). b) Amplificação específica a partir da sonda para o subtipo G para a amostra 251 (ROX). São ainda representados a curva representativa dos controlos negativos de reacção.

3.5 Confirmação da identidade genética das estirpes virais presentes em amostras clínicas: caracterização de cada região do genoma por análise filogenética

3.5.1 Amplificação, purificação e sequenciação de produtos de PCR

Para a avaliação da natureza genética das estirpes virais presentes nas amostras clínicas estudadas, e tendo em consideração o tamanho reduzido dos fragmentos universais específicos (referidos nas secções anteriores), a análise de sequências virais teve por base a obtenção de novos produtos de amplificação. Para tal foram efectuadas novas reacções de PCR utilizando-se, desta feita, uma abordagem de *heminested-PCR*. Esta baseou-se na utilização de um *primer* universal externo em conjunto com um *primer* universal interno, facto este que permitiu a amplificação indiscriminada de sequências virais a partir de todas as amostras, independentemente da sua identidade genética. Como tal esta abordagem originou fragmentos que não se restringiram às regiões alvo do genoma viral estudadas (fragmentos específicos amplificados ao longo deste trabalho). Assim, e apenas como exemplo, para a região Pr foi analisado um

fragmento que resultou da amplificação por *heminested*-PCR com uma dimensão total de 635pb incluindo parte da região da RT, ao passo que o fragmento específico obtido nos ensaios de amplificação específica apresentava cerca de 108pb. A dimensão, regiões do genoma abrangentes dos fragmentos esperados (por região estudada), obtido para cada uma das amostras clínicas analisadas, encontra-se representado na Tabela 3.8. As reacções de amplificação foram conduzidas de forma semelhante às reacções de amplificação universal descritas anteriormente (temperatura de hibridação de 55°C) a partir de 1µl de uma diluição a 1:1000 das reacções contendo amplicões universais externos (visíveis em gel de agarose) ou 5µl das mesmas quando a sua análise em gel de agarose não revelou qualquer produto.

Tabela 3.8 – Dimensões dos fragmentos esperados para cada uma das regiões analisadas. Estão ainda representadas as regiões comportadas pelos fragmentos amplificados.

Região estudadas	Gag	Pr	RT	Rev	Gp120	Gp41
Região alvo abrangentes	p24Gag-Protease	Protease-p51RT	p51RT-p51RT	Vif-Rev	Vpu-Gp120	Gp41-Nef
Tamanho do Fragmento esperado (≈pb)	540	640	570	670	600	680

Uma vez amplificados os produtos de PCR foram submetidos a uma etapa de purificação, tal como descrito no capítulo 2 (secção 2.9) deste trabalho. Os produtos purificados foram sequenciados pelos serviços de sequenciação da empresa STAB VIDA (Portugal), recorrendo à técnica de terminação da polimerização em cadeia inicialmente descrita por Sanger. A sequenciação foi efectuada apenas num dos sentidos usando um dos *primers* da amplificação universal anterior (*heminested*-PCR), o *primer* que apresentasse menor números de bases degeneradas para cada região, respectivamente.

3.5.2 Análise filogenética das sequências virais

A identificação do subtipo referente às sequências virais de HIV-1 amplificadas a partir das amostras clínicas incluídas neste trabalho foi efectuada com base na análise das relações filogenéticas entre as sequências nucleotídicas obtidas e um extenso conjunto de outras ditas de referência. Para tal, as sequências obtidas foram alinhadas juntamente com sequências representativas de cada um dos diferentes subtipos e algumas CRF de HIV-1 com origens geográficas distintas (África do Sul - **ZA**, Botswana - **BW**, Brasil - **BR**, Etiópia - **ET**, Índia - **IN**, Quénia - **KE**, Uganda - **UG**, França - **FR**, Espanha - **ES**, Portugal - **PT**, Estados Unidos da América - **US**, Cuba - **CU**, Gana - **GH**, Senegal - **SN**, República Democrática do Congo - **CD**, Camarões - **CM**, Suécia - **SE**, Burkina Faso - **BF**, Chipre - **CY**, República Centro-Africana - **CF**, Rússia - **RU**, Coreia do Sul - **KR**, China - **CN**, Chade - **TD**, Holanda - **NL**, Alemanha - **DE**, Geórgia - **GE**, Nigéria - **NG**, Itália - **IT**, Dinamarca - **DK**, Bélgica - **BE**, Uzbequistão - **UZ**, Emirados Árabes Unidos - **AE**) disponíveis em bases de dados de acesso público. Com base neste alinhamento múltiplo foi construída uma árvore filogenética por cada região do genoma viral estudada. A análise das diferentes árvores filogenéticas obtidas revelou que as sequências obtidas neste trabalho surgiram agrupadas com as referências dos subtipos B e G, e das formas genéticas CRF02_AG e CRF14_BG. Foi disto excepção a amostra 317 que surgiu agrupada com referências do subtipo D para a região de Gag, e da amostra 658 que surgiu agrupada com referências sub-subtipo F1 para a região Pr (Figura 3.12). Para algumas das sequências virais em análise, nomeadamente das amplificadas a partir das amostras clínicas 317 e 471 (região de Gag) e das amostras 107 e 658 (região Pr), verificou-se divergência ou uma projecção evidente das mesmas como nós terminais de ramos longos, relativamente a todas as restantes sequências nucleotídicas incluídas no mesmo grupo monofilético. Este facto pode reflectir a ocorrência de eventos de recombinação entre sequências virais de subtipos distintos, a presença de estirpes virais divergentes ou, ainda, a ocorrência sistemática de hiperadenilação ao longo das sequências virais analisadas. Uma vez que a primeira das hipóteses avançadas é, frequentemente, a responsável pelo fenómeno observado, as sequências virais em questão foram ainda inspeccionadas de forma a verificar o seu possível carácter recombinante (tal como descrito em seguida na

secção 3.5.4). Os resultados obtidos através da sequenciação e análise filogenética encontram-se resumidos no anexo 1. Até à data de conclusão deste manuscrito, e por razões que se prendem com as limitações de tempo impostas, não foi possível levar a cabo a sequenciação e análise filogenética de todas as sequências virais amplificadas de todas as amostras clínicas estudadas.

A informação recolhida através das árvores filogenéticas e uma vez que estas análises fornecem informações tão robustas quanto possível sobre as relações evolutivas entre as diferentes sequências virais vieram complementar os resultados obtidos por ferramentas de genotipagem automática como a *REGA* versão 2.0 e pela ferramenta *Blast*. Embora esta última não apresente qualquer inferência de modelos de evolução, a sequência em questão é comparada com as sequências disponíveis na base de dados. Os dados obtidos coincidiram, essencialmente, com os resultados obtidos através da análise filogenética.

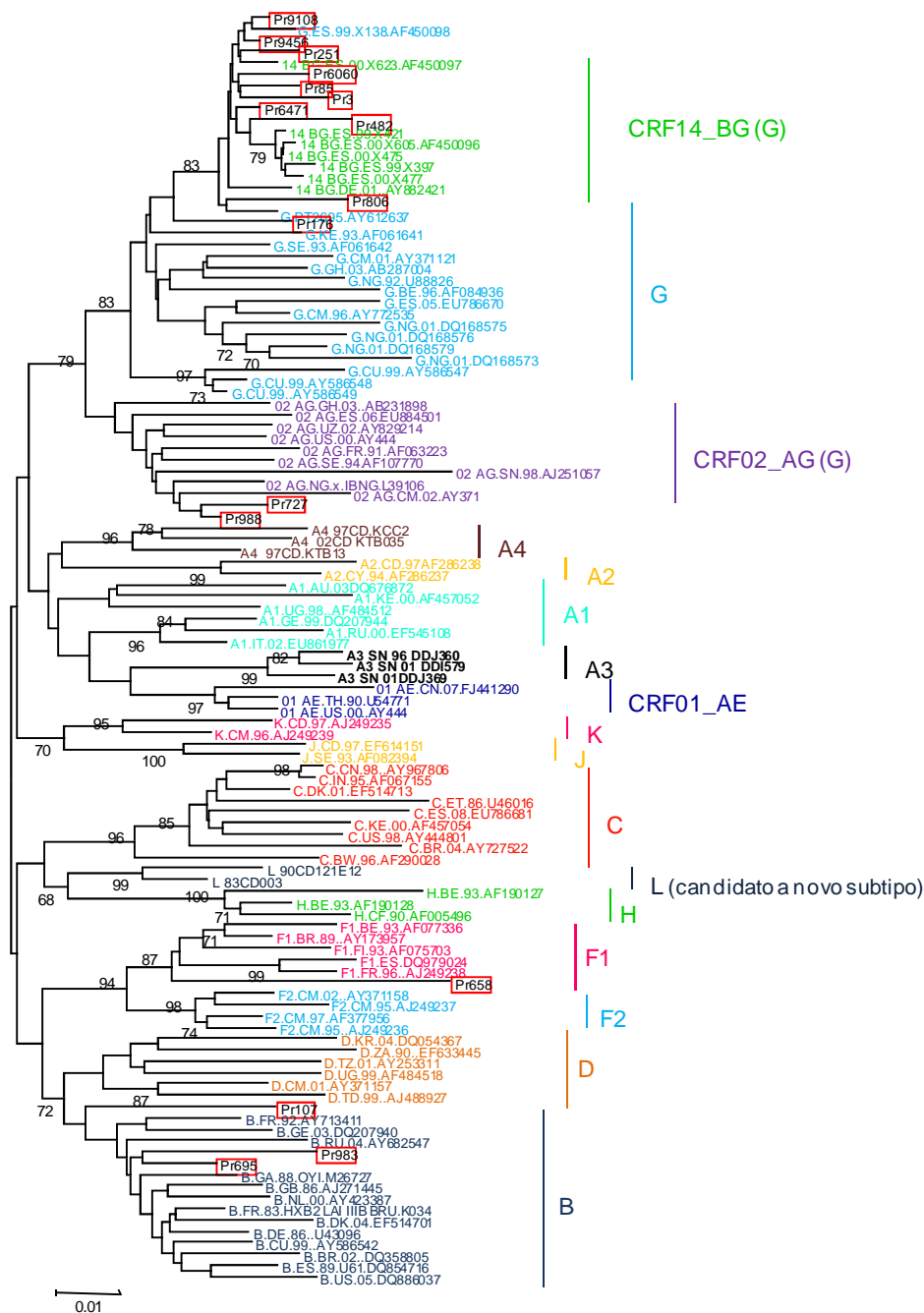


Figura 3.12 – Relações filogenéticas entre estirpes de HIV-1 baseadas na análise das sequências nucleotídicas da região codificante da protease viral. A árvore filogenética representa as relações filogenéticas entre as sequências da região da Protease das estirpes de HIV-1 obtidas e estirpes virais de referência (grupos M). As estirpes pertencentes aos indivíduos estudados neste trabalho estão assinaladas pelo contorno vermelho, apresentado ainda o prefixo Pr acompanhado pelo número da amostra clínica de

onde foram amplificadas. As restantes estirpes constituem as estirpes de referência identificadas pelo subtipo, país de origem, ano e número de acesso. A árvore foi obtida pelo método de *neighbor-joining*, a partir de matrizes de distâncias calculadas pelo método *Kimura* a dois parâmetros. Para alguns ramos estão ainda indicados os valores de *bootstrap* (percentagem de 1000 réplicas) acima de 70%. A escala representa a distância genética em número de substituições nucleotídicas por sítio.

3.5.3 Análise comparada dos resultados de genotipagem baseados em PCR com *primers* específicos e da análise filogenética das sequências estudadas

A informação recolhida através dos vários procedimentos analíticos realizados ao longo deste trabalho, tendo em vista a genotipagem de HIV-1 no universo de amostras clínicas estudadas foi analisada de forma comparativa. Pela inexistência de informação integral relativa ao genoma completo de cada uma das estirpes clínicas em estudo, optou-se por referir os dados disponíveis resultantes de trabalhos anteriores em que se procedeu à análise de sequências virais amplificadas das mesmas amostras biológicas. Estes dados diziam respeito, essencialmente, à análise por HMA de parte das regiões Gag (p24-p7) e Env (C2-V3-C3) ou por análise filogenética de sequências virais correspondentes a secções das regiões que codificam a protease e a transcriptase reversa virais, assim como da região regulatória LTR do extremo 5' do genoma proviral (78, 30, 29). É importante salientar que as análises parciais efectuadas anteriormente, embora tenham origem na mesma amostra biológica, na maioria dos casos, as regiões analisadas e o material biológico de origem (RNA viral vs. DNA proviral) foi distinto.

Para a região Gag, dos 29 fragmentos comportando sequências virais amplificadas das amostras analisadas em trabalhos anteriores por HMA, para 8 foram registados resultados de amplificação inespecíficos na genotipagem efectuada com base da abordagem por PCR/gel de agarose convencional. Em alternativa, a genotipagem baseada na utilização do PCR em tempo-real revelou 6 amplificações inespecíficas. Para ambas as estratégias, 5 dos resultados inespecíficos de genotipagem (317, PG81, PG176, PG178, PG184, PG187) correspondiam a amostras para as quais as análises realizadas previamente identificaram as sequências em causa como pertencentes ao subtipo D (317) e subtipo C (todas as outras). As restantes estavam identificadas para o subtipo B (073) e G (251). Por seu lado a sequenciação e análise filogenética dos fragmentos que incluíam a região alvo dos *primers* específicos para esta região do

genoma revelou que duas das onze amostras estudadas apresentavam discordância no resultado obtido, comparativamente às amplificações efectuadas (PCR convencional e PCR em tempo-real). Por um lado sequências virais da amostra 508 foram amplificadas especificamente com o *primer* discriminativo de sequências do subtipo G, mas a sequência analisada foi agrupada com sequências incluídas no agrupamento monofilético contendo referências do subtipo B. Por sua vez as sequências virais da amostra 176 foram amplificadas especificamente pelo *primer* discriminativo de sequências do subtipo B, enquanto que, a sua análise filogenética revelou o seu agrupamento com as amostras de referência para o subtipo G. As sequências virais da amostra 709 foram identificadas por análise filogenética, PCR/gel agarose convencional/tempo-real como pertencentes ao subtipo B, contrariamente ao resultado obtido por e análise prévia por HMA, que concordaram na atribuição do subtipo G às sequências em causa.

Em consequência dos resultados discordantes obtidos, optámos por uma avaliação da potencial natureza mosaico destas sequências por uma análise de eventos de recombinação através da técnica de *bootscanning* (secção 3.5.4).

No que diz respeito à análise das sequências virais que incluíram a região Pr foi registado um aumento de resultados discriminatórios na ordem dos 18% quando a estratégia de genotipagem baseada em PCR em tempo-real foi comparada à abordagem por PCR/gel de agarose convencional, esta última tendo revelado apenas 36 amplificações específicas (subtipo B ou G). Independentemente da abordagem utilizada, foram registadas amplificações inespecíficas de sequências virais encontradas nas amostras para as quais uma análise prévia tinha revelado a presença de vírus do subtipo C (PG81, PG176, PG178, PG184, PG187). A análise de sequências virais provenientes das restantes amostras revelou um resultado de amplificação discriminatório para penas um dos subtipos por PCR em tempo-real. O cruzamento de dados das amplificações específicas efectuadas por recurso ao PCR em tempo-real e da informação (análise filogenética) para a totalidade da região Pr (297pb), previamente disponível para um conjunto de 10 amostras (2, 3, 13, 16, 24, 38, 72, 85, 98, 107), revelou que todas as amostras evidenciaram um resultado coincidente em relação à identificação do subtipo viral. Os resultados de genotipagem baseados nas amplificações por PCR em tempo-real foram ainda comparados com os resultantes de análise filogenética das sequências de

amplificações obtidos como descrito em 3.5.1, amplificados a partir de 16 amostras clínicas. Esta comparação demonstrou uma coincidência na identificação do subtipo genético para a maioria das sequências analisadas; contudo, para duas amostras foram observadas divergências nas genotipagens efectuadas. A sequência da amostra 658 foi amplificada de uma forma discriminativa baseada na utilização do *primer* específico para sequências do subtipo G (PCR/ gel de agarose e PCR em tempo-real) ao passo que a análise filogenética revelou a segregação das sequências analisadas juntamente com as referências do sub-subtipo F1. No caso da amostra 107 embora os resultados da amplificação das sequências por PCR/gel de agarose convencional, PCR em tempo-real e análise filogenética exclusivamente da região Pr (esta última recolhida em outros trabalhos) tenham sido concordantes, identificando a amostra como pertencente ao subtipo G, na análise das relações filogenéticas do fragmento de 635pb obtido por *heminested-PCR* surgiu agrupada juntamente com as sequências de referência do subtipo B.

Para a região RT a abordagem por PCR em tempo-real permitiu discriminar duas das 7 amostras para as quais, por PCR/gel de agarose convencional, foram obtidas amplificações de sequências virais baseadas na utilização de mais do que um *primer* específico. Fragmentos do genoma viral das estirpes presentes nas amostras 8731 e PG176 foram amplificados especificamente (discriminadas) por PCR em tempo-real quando não tinham sido visualizados produtos de amplificação em gel de agarose. Para 8 amostras do conjunto de 10 para as quais estava disponível informação prévia (inferência filogenética), os resultados obtidos pelas duas abordagens de PCR utilizadas foram concordantes. A amostra 13 deu origem a um sinal de amplificação não discriminatório quando utilizados os *primers* específicos para sequências virais do subtipo G e da forma genética CRF02_AG, embora tenha agrupado com as amostras de referência do subtipo G por análise filogenética. Apesar da concordância (subtipo G) revelada pela comparação dos resultados de amplificação específica por PCR/gel de agarose convencional e por PCR em tempo-real de sequências virais da amostra 38, a análise das relações filogenéticas realizada previamente em outros trabalhos identificou-a como pertencente ao subtipo B (com valores de *bootstrap* superiores a 70%). Relativamente às amostras identificadas previamente para o subtipo C, quando amplificadas por PCR/gel de agarose convencional 4 amostras das 5 analisadas

apresentaram um resultado inespecífico, sendo que para a amostra PG176 não foi detectado qualquer produto amplificado. Por seu lado, a partir do ensaio em PCR em tempo-real 3 das amostras revelaram um resultado de amplificação específico para CRF02_AG (PG176, PG184, PG187), ao passo que as restantes evidenciaram um perfil de amplificação duplo para o subtipo G e CRF02_AG. Para o grupo de 8 amostras sequenciadas e analisadas filogeneticamente neste trabalho, os resultados das amplificações específicas foram concordantes, excepto para a uma amostra. A amostra em questão (508) apresentou um resultado duplo para ambos os subtipos B e G (independentemente da abordagem utilizada), enquanto que a genotipagem de sequências virais baseada na análise filogenética agrupou-as com as referências do subtipo B.

Para a região Rev, das 5 amostras (72, 114, 806, 983, 6060) que deram origem a resultados de amplificação dupla na abordagem por PCR/gel de agarose convencional, apenas a genotipagem de sequências virais da amostra 72 não foi discriminativa por PCR em tempo-real, tendo sido registada uma amplificação dupla independentemente da abordagem utilizada. Por outro lado, em particular no que diz respeito às sequências amplificadas da amostra 3013, embora a genotipagem baseada na PCR/gel de agarose convencional tenha revelado um produto de amplificação mediante a utilização exclusiva do *primer* específico para o subtipo B, esta mesma amostra revelou um resultado duplo de amplificação em PCR em tempo-real. A abordagem por PCR em tempo-real permitiu ainda a amplificação de sequências virais da amostra 919 (específica para o subtipo B) para a qual não se detectou previamente amplificação na abordagem PCR/gel de agarose convencional. Para o grupo de amostras genericamente designadas com o prefixo PG, apenas para uma (PG81) foi registado um perfil de amplificação específico para o subtipo G, independentemente da abordagem utilizada, sendo que para as restantes não foram detectados produtos amplificados. Para o conjunto de 5 amostras (695, 727, 806, 983, 988) sequenciadas e analisadas filogeneticamente para esta região do genoma viral verificou-se uma concordância de resultados com os dados obtidos nas amplificações específicas para todas as amostras analisadas.

A região de Gp120 foi a região por entre as estudadas para a qual um maior número de amostras analisadas revelou resultados inespecíficos de amplificação,

independentemente da estratégia experimental utilizada. A genotipagem baseada na amplificação de sequências por PCR/gel de agarose convencional revelou resultados triplos de amplificação para as amostras 727, 988, 9456 e duplos para as amostras 271, 508, 695 e 983. No entanto, para as amostras que apresentaram um resultado duplo, foi possível amplificar fragmentos virais específicos para apenas um dos subtipos por recurso ao PCR em tempo-real, resultados, estes, coincidentes com subtipagem previamente efectuada por HMA. Em relação às restantes 30 amostras para as quais foi recolhida previamente informação relativa à identificação do subtipo viral (por HMA), a maioria dos resultados obtidos neste trabalho (utilizando PCR em tempo-real) foram com ela coincidentes. No entanto, para um conjunto de amostras registaram-se algumas discordâncias. Para todas as amostras apresentando estirpes virais identificadas incluídas no subtipo C foram obtidas amplificações inespecíficas com base em ambas as estratégias experimentais utilizadas. Para a amostra 13 apesar de não ter sido detectado produto de amplificação visível em gel de agarose, por PCR em tempo-real foi registada uma amplificação com perfil duplo para os subtipos B e G. A juntar a estas, as amostras 4132 e 4773 estiveram na origem de resultados de amplificação dupla (subtipo G e CRF02_AG). Finalmente, a amostra 471 apresentou um perfil de amplificação específica para o subtipo B, apesar de uma análise prévia por HMA ter caracterizado sequências virais como pertencentes ao subtipo A. A informação fornecida pela subtipagem por HMA das sequências pertencentes às amostras 4132 e 4773 coincidiu com o resultado obtido através da análise filogenética, no entanto, estas sequências foram ainda avaliadas por análise de recombinantes (ver secção 3.5.4). As sequências virais da amostra 317 às quais foi atribuído o subtipo D (HMA) foi amplificada especificamente pelos *primers* específicos para o subtipo B. No caso das sequências virais das amostras 955 e 9000 identificadas para o subtipo F (HMA), esta informação foi discordante com os resultados obtidos através da amplificação específica das sequências virais independentemente da estratégia utilizada, a primeira foi amplificada para o subtipo B e a segunda duplamente com os *primers* específicos para CRF02_AG e G. As sequências virais das amostras 8731 e 9108, identificadas como pertencentes ao subtipo G por HMA, originaram amplificações específicas com o *primer* para o subtipo B, independentemente da estratégia utilizada. Para as restantes amostras analisadas

filogeneticamente (271, 508, 695, 836, 983) verificou-se uma concordância entre os resultados obtidos através da amplificação específica e a inferência filogenética.

Para a região Gp41 das 50 sequências virais pertencentes às amostras testadas, apenas duas (695 e 806) apresentaram um resultado duplo de amplificação por PCR convencional, as quais foram discriminadas no ensaio de PCR em tempo-real (695 para o subtipo B, 806 para o subtipo G). A sequência pertencente à amostra 695 para esta região foi ainda analisada filogeneticamente, a qual suportou o resultado obtido por PCR em tempo-real. Para além desta foi ainda analisada por inferência filogenética a sequência obtida a partir da amostra 1544, que surgiu agrupada com as sequências do subtipo G, estando de acordo com o resultado obtido através das amplificações por PCR em tempo-real. Por outro lado, para as sequências virais pertencentes às amostras 72, 176, 919 não foram detectados produtos amplificados visíveis em gel de agarose mas foram amplificados, especificamente, para um dos subtipos B ou G por PCR em tempo-real. Para a amostra 271 não foram detectadas amplificações de sequências virais independentemente da abordagem utilizada e o mesmo se verificou para o conjunto de amostras PG81, PG176, PG178, PG184, PG187. As restantes amostras apresentaram resultados coincidentes entre as estratégias utilizadas.

Para esta mesma região foram ainda comparados os dados dos ensaios efectuados por PCR em tempo-real mas recorrendo a sondas do tipo *Taqman* para 27 amostras escolhidas aleatoriamente. Para 14 (52%) as amostras os resultados obtidos foram concordantes com as observações efectuadas anteriormente, contudo para as restantes amostras não foram detectadas curvas fluorescentes relativamente a possíveis amplificações. Os resultados desta análise encontram-se resumidos no anexo 1.

3.5.4 Avaliação, por *bootscanning*, da natureza recombinante de algumas das sequências de HIV-1 analisadas

Para as sequências virais que nas árvores filogenéticas construídas se apresentavam como nós terminais de ramos com projecção evidente relativamente aos demais, ou que, em alternativa, estiveram na base de discordâncias registadas entre as análises efectuadas tendo em vista a sua caracterização (genotipagem baseada na

amplificação específica por PCR, sequenciação/análise filogenética e HMA) procedeu-se à avaliação da natureza recombinante das mesmas por recurso à técnica de *bootscanning*.

A análise da região Gag do genoma de HIV-1 amplificada da amostra 317 e agrupada com as amostras do subtipo D por análise filogenética, evidenciou a sua natureza mosaico. De facto, a análise por *bootscanning* efectuada confirmou a sua clara classificação entre o subtipo D, mas falhou na identificação da sua origem genética na extremidade 3' (Figura 3.13a). Em relação à amostra 471, cujas sequências virais na região Gag se demonstrou agruparem com as amostras de referência do subtipo G (segregando com sequências de tipo CRF14_BG), a análise de *bootscanning* não revelou quaisquer evidências óbvias de recombinação intergenotípica (embora para alguns pontos específicos da sequência nucleotídica tenha reflectido alguma divergência em relação as estirpes de referência utilizadas para o CRF14_BG, Figura 3.13b). Para as sequências virais pertencentes à amostra 709, embora estas tenham apresentado resultados concordantes entre as amplificações específicas por PCR/gel de agarose convencional, PCR em tempo-real e análise filogenética, identificando-a como pertencente ao subtipo B, a análise prévia por HMA identificou-a para o subtipo B. O gráfico de *bootscanning* relativo à análise do fragmento amplificado que cobre esta região em particular revelou, no que a esta amostra diz respeito, uma grande proximidade filogenética com as sequências de referência do subtipo B utilizadas neste trabalho (Figura 3.13c). Para a amostra 176 e embora as sequências virais da mesma tenham originado amplificações específicas por PCR/ gel de agarose convencional e PCR em tempo-real para o subtipo B, o resultado obtido a partir da análise filogenética do fragmento obtido por *heminested-PCR* identificou-a como pertencente ao subtipo G. O gráfico de *bootscanning* relativo à análise do fragmento amplificado evidenciou um fenómeno de recombinação entre o subtipo B (extremidade 5') e sequências G filogeneticamente semelhantes à sequência G que deverá ter estado na origem de forma recombinante CRF14_BG (extremidade 5') (Figura 3.13d). Ainda relativamente às sequências para a região de Gag amplificadas da amostra 508, embora a utilização de *primers* específicos e a análise prévia por HMA tenham concordado na sua associação ao subtipo G, a análise filogenética identificou as sequências em causa como pertencente ao subtipo B. Contudo, a análise por *bootscanning* desta mesma sequência

falhou na atribuição clara de um subtipo viral à região central da mesma, ainda que na sua maior parte esta deva compreender uma origem no subtipo B (Figura 3.13e).

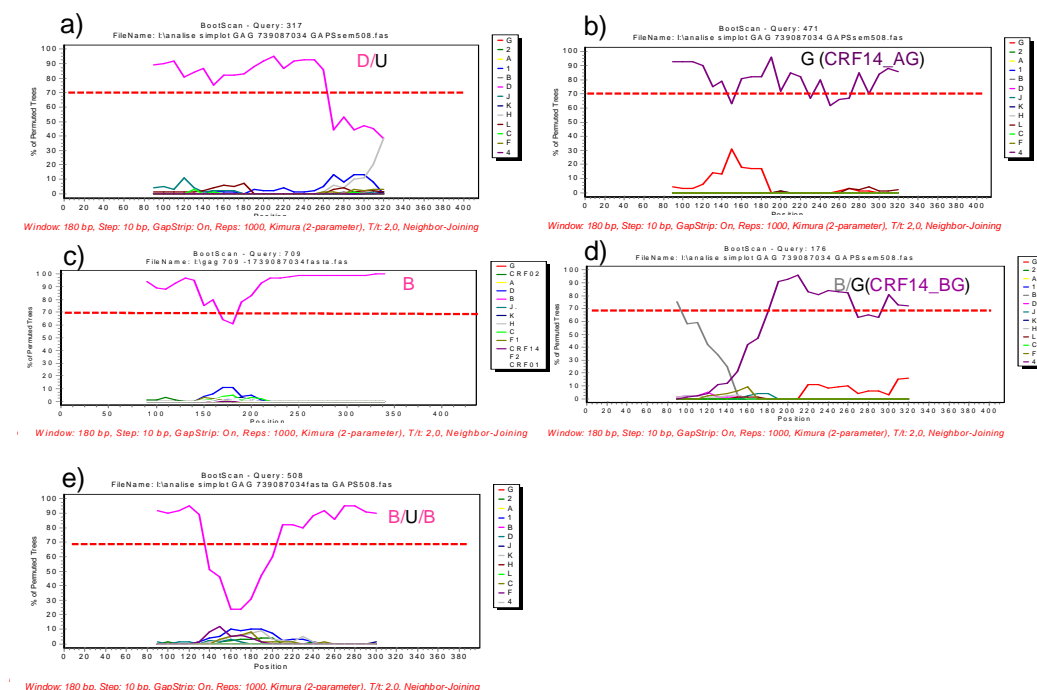


Figura 3.13 – Análise da potencial natureza recombinante de sequências virais (região Gag). Resultados obtidos pela análise por *bootscanning* para as sequências amplificadas da amostra 317 (a), 471 (b) e 709 (c), 176 (d), 508 (e). Nestas análises consideraram-se significativos os resultados acima dos 70%.

A análise das sequências virais correspondentes à região Pr, amplificadas especificamente com *primers* para o subtipo G a partir da amostra 658 revelou, por análise filogenética, o seu agrupamento com as sequências do sub-subtipo F1. Contudo, esta mesma sequência apresentou-se no extremo de um ramo evidentemente prolongado relativamente às restantes sequências do agrupamento. A avaliação da sua potencial natureza recombinante evidenciou dois possíveis locais de *crossing-over* que separam a secção da sequência com clara homologia com as referências do sub-subtipo F1, de uma secção não subtipável por ausência de semelhança com as sequências de referência utilizadas (Figura 3.14a). Para a amostra 107, as amplificações específicas da sequência viral (PCR/gel agarose convencional, PCR em tempo-real) e a informação prévia da análise filogenética exclusivamente do fragmento correspondente à região Pr (297pb)

identificaram a sequência viral dela amplificada como pertencente ao subtipo G. No entanto, a análise filogenética de um fragmento de maiores dimensões (635pb, como referido em 3.5.1) identificou a amostra como pertencente ao subtipo B, apesar de ser a primeira sequência a divergir em relação ao grupo monofilético. O resultado da análise de recombinantes (*bootscanning*) evidenciou um evento de recombinação intergenotípica caracterizado pela existência de um região (extremidade 5') não subtipável por ausência de semelhança com as sequências de referência utilizadas, ao passo que a restante sequência apresentou-se filogeneticamente aparentado com as amostras de referência do subtipo B (Figura 3.14b).

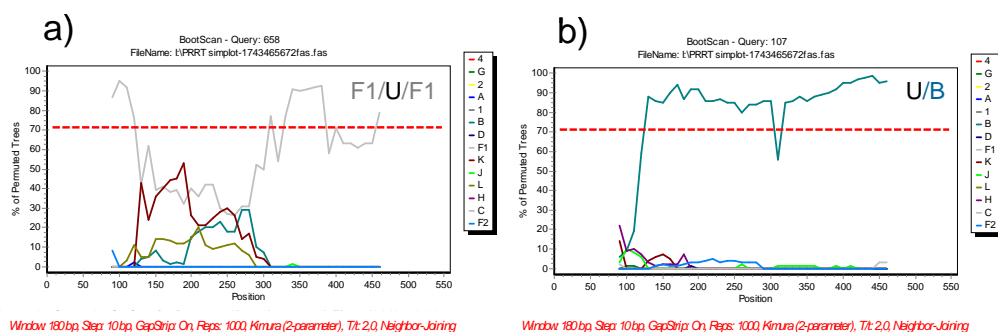


Figura 3.14 – Análise da potencial natureza recombinante de sequências virais (região Pr). Resultados obtido pela análise por *bootscanning* para as sequências amplificadas da amostra 658(a), 107 (b). Nestas análises consideraram-se significativos os resultados acima dos 70%.

Para a região RT, as sequências virais presentes na amostra 508 originaram amplificações de forma dupla com ambos os *primers* específicos para os subtipos B e G e que formaram um agrupamento com as amostras de referência do subtipo B através da análise filogenética. Como tal, procedeu-se a inspeção da sua potencial natureza recombinante por *bootscanning* mas esta não revelou quaisquer indicações que sugerissem que eventos de recombinação tenham estado na sua origem (Figura 3.15).

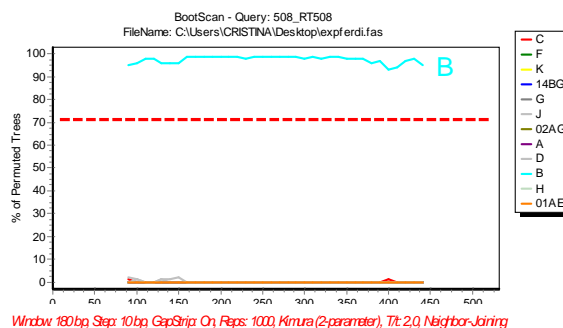


Figura 3.15 – Análise de bootscanning da região RT das sequências virais amplificadas da amostra 508. Nestas análises consideraram-se significativos os resultados acima dos 70%.

Para a região Gp120 foram analisadas duas sequências virais amplificadas das amostras 4132 e 4773, ambas apresentando um perfil de amplificação duplo com os *primers* específicos para sequências dos subtipo G e da forma CRF02_AG. A sua análise filogenética indicou a sua segregação com as amostras de referência pertencentes à CRF02_AG, resultado este que veio a ser confirmado através da inspeção do seu potencial recombinante por *bootscanning* (Figura 3.16).

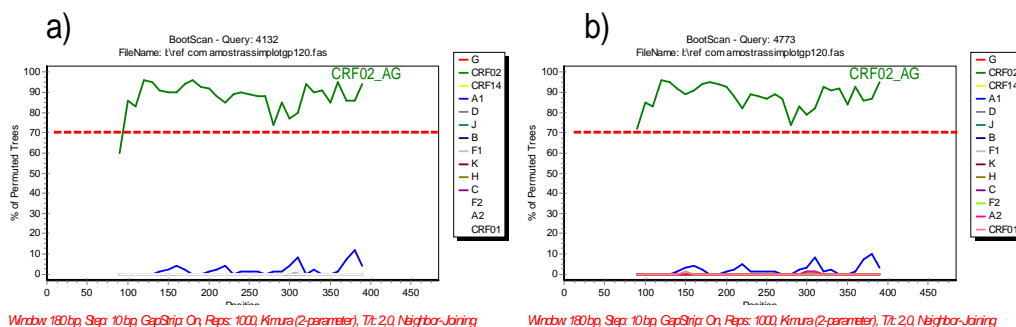


Figura 3.16 – Análise da potencial natureza recombinante de sequências virais (região Gp120). Resultados obtido pela análise por *bootscanning* para as sequências amplificadas da amostra 4132 (a), 4773 (b). Nestas análises consideraram-se significativos os resultados acima dos 70%.

4. Discussão e Conclusões

4. Discussão e Conclusões

A pandemia da qual o HIV-1 é responsável caracteriza-se, entre outros, pela circulação simultânea de múltiplas variantes virais, classificadas, do ponto de vista genético, em pelo menos 9 subtipos distintos e numa multiplicidade de formas recombinantes (8, 106). A diversidade genética do HIV-1 apresenta-se, por si só, como um enorme desafio face à tentativa de monitorização (tendo em conta a sua caracterização genética) das estirpes circulantes em diferentes áreas do Globo. Para além de permitir avaliar, em termos qualita/quantitativamente, a distribuição dos subtipos circulantes na maioria dos países, realça-se a importância de que esta monitorização permite ainda analisar a dinâmica da evolução viral. Esta pode ser traduzida na identificação dos eventos de recombinação que dão origem aos diferentes vírus com genoma mosaico e dos factores que contribuem para a própria recombinação (47), ou mesmo na associação de estirpes virais a comportamentos de risco (57). Por outro lado, a descrição precisa dos subtipos e formas virais recombinantes circulantes reveste-se de particular importância em regiões onde se pretendam implementar ensaios de potenciais vacinas (47). Para tal, são necessárias estratégias que permitam a análise de um elevado número de amostras em tempo útil, de modo a que a vacina candidata contemple as estirpes responsáveis pela epidemia numa dada região (119). Na sua maioria, as ferramentas disponíveis para a caracterização genética do HIV-1 (HMA, sequenciação de pequenos fragmentos ou, em alternativa, a sequenciação do genoma completo, entre outros) revelaram-se de pouca utilidade tendo em vista este tipo de intervenções (47). Tal facto, por si só, fomentou o desenvolvimento de estratégias experimentais alternativas que permitam caracterizar grandes números de estirpes virais. É neste contexto que os ensaios de MHA surgem como uma abordagem experimental a considerar.

Estes ensaios (MHA) permitem, simultaneamente, o rastreio de um elevado número de amostras num curto espaço de tempo, com custos relativamente reduzidos ao mesmo tempo que fornecem informação genética suficiente para permitir a caracterização, do ponto de vista genético, e com uma exactidão relativamente elevada, das várias estirpes virais circulantes numa qualquer região do Globo (47, 56). Embora na sua génese todos os ensaios deste tipo descritos na literatura tenham sido delineados

de forma semelhante, isto é, seguindo princípios técnicos comuns a todos eles, contudo, normalmente desenvolvidos/adaptados a nível regional, uma vez que as estirpes circulantes numa dada região poderão ser distintas das que circulam noutras. Por outras palavras, de modo a implementar, correctamente, um ensaio de MHA para uma dada região, é necessário um conhecimento prévio sobre quais os subtipos virais e formas recombinantes que dominam a epidemia em questão (46). Em Portugal, como referido ao longo deste trabalho, os vários estudos de caracterização genética efectuados, apesar de, na sua maioria, fornecerem apenas informação parcial sobre as características do genoma das estirpes virais estudadas, demonstraram uma elevada prevalência de vírus de subtipos não-B, sendo dominada pelos subtipos B e G e as CRF02_AG e CRF14_BG (30, 75, 76, 78). Tal facto foi, em grande medida, condicionado pelos estreitos laços históricos que Portugal manteve com países africanos. Deste modo, pretendeu-se com este trabalho desenvolver um ensaio de tipo MHA tendo em vista uma caracterização genética, tão célere quanto possível, das estirpes de HIV-1 dominantes na epidemia Portuguesa. O desenvolvimento deste ensaio e a sua aplicação em Portugal deverá permitir identificar rapidamente um elevado número de estirpes a custos reduzidos, sobretudo quando comparada, como alternativa, à sequenciação completo do genoma das estirpes de HIV-1 encontradas.

Uma vez que são analisadas, em simultâneo, várias regiões do genoma viral, a concepção adequada deste tipo de ensaios reveste-se de uma particular importância, tendo em conta as características da epidemia Portuguesa, onde cerca de 43% das infecções por HIV-1 estão associadas a UDI (14). Como já foi referido anteriormente, este tipo de epidemia, associada a grupos populacionais com comportamentos de risco, tem um enorme impacto na evolução e diversidade genética do vírus, uma vez que potencia a ocorrência de co-infecções, abre espaço à possibilidade de ocorrência de eventos de recombinação e, consequentemente, de aparecimento de novos vírus recombinantes.

Na sua formulação clássica os ensaios de MHA integram uma primeira amplificação, *in vitro*, de sequências nucleotídicas virais mediante a utilização de *primers* universais, seguida de uma segunda amplificação em que o produto da reacção anterior é utilizado como matriz. Estes ensaios são realizados (pelo menos no que à segunda etapa diz respeito) num termociclador que permita, em tempo-real, seguir o

processo de amplificação, durante o qual a identificação selectiva de sequências virais resulta da presença, nas segundas misturas de reacção, e em simultâneo com *primers* universais, de sondas subtipo-específicas (46, 56, 57, 119). Esta formulação define o MHA “convencional”. As variantes experimentais alternativas, apresentadas ao longo deste trabalho, foram desenvolvidas de forma a maximizar o sucesso dos processos de amplificação, considerando que este pode ser largamente determinado pela baixa carga viral de algumas amostras biológicas (por vezes consequência de uma má preservação das amostras em questão, ou da sua exposição a ciclos repetidos de congelação/dcongelação) (115). Nas variantes aqui propostas, em alternativa à abordagem “convencional” do MHA, a amplificação, por PCR, de sequências de HIV-1 pode ser registada através da visualização dos produtos obtidos em gel de agarose ou, em alternativa, pela detecção de curvas de amplificação específica num formato de PCR em tempo-real realizado na presença de *SYBR[®] Green I*. O processo de amplificação aqui proposto baseia-se na obtenção de fragmentos de genoma viral em duas ou três etapas sucessivas, numa combinação de estratégias de *nested* e *heminested*-PCR. Apresenta-se, ainda, como uma alternativa à abordagem “convencional” do MHA passível de ser implementada num contexto financeiramente limitado.

Em termos gerais, o sucesso do MHA está dependente do delineamento dos *primers* de amplificação universal e específica, os quais são desenhados a partir de um alinhamento múltiplo de sequências, idealmente representativo da diversidade genética viral registada na região para qual se pretende implementar o método (56, 57). A escolha das sequências virais para a construção deste alinhamento múltiplo teve como base a informação recolhida na bibliografia sobre as características da epidemia do HIV-1 em Portugal, em particular, a informação sobre quais os subtipos e formas genéticas recombinantes circulantes mais frequentemente encontradas (30, 75, 76, 78). Neste alinhamento foram incluídas, preferencialmente, e sempre que disponíveis nas bases de dados de acesso livre, sequências virais pertencentes a estirpes de HIV-1 identificadas em Portugal e/ou países vizinhos (desde Espanha a outros países europeus), com o intuito de que este alinhamento abrangesse, na medida do possível, a diversidade genética viral num perímetro geográfico definido. Assim sendo, o alinhamento construído teve por base um conjunto de 57 sequências virais correspondendo a genomas completos de estirpes de referência de HIV-1, tendo-se, e tal

como acima referido, dado particular ênfase às quatro variantes dominantes da epidemia Portuguesa. Foram, assim, alinhadas 19 sequências correspondentes a genomas de vírus do subtipo B, 12 a vírus do subtipo G, 11 da CRF02_AG e 7 da CRF14_BG. As restantes corresponderam a duas sequências virais classificadas em cada um dos subtipos A e F, e uma para cada um dos subtipos C, H, J e K.

Na análise de um alinhamento múltiplo de sequências que tenha em vista o delineamento de *primers* para um ensaio de MHA, e que inclua sequências de vírus recombinantes, deverá ser dada especial atenção à estrutura mosaico do genoma dos mesmos. Neste contexto, torna-se particularmente importante identificar, tão correctamente quanto possível, os pontos de recombinação que caracterizam estes genomas virais (no caso em particular, as CRF02_AG e CRF14_BG). Da análise do alinhamento construído resultou a identificação de 7 regiões distintas (Gag, Pr, RT, In, Rev, Gp120, Gp41), as quais serviram de alvos para amplificação específica, e identificação inequívoca, da origem genética das estirpes analisadas. Estas regiões caracterizaram-se, simultaneamente, por serem altamente conservadas na sua totalidade (ou quase totalidade) quando as sequências de vírus de um mesmo subtipo/CRF foram comparadas entre si, mas apresentaram-se distintas quando, em alternativa, se comparavam sequências de vírus classificados em subtipos/CRF distintos. A flanquear estas regiões reconheceram-se outras altamente conservadas por entre todos os subtipos do HIV-1, as quais foram utilizadas para a definição dos *primers* de amplificação universal (externos e internos). No processo de delineamento da sequência dos *primers* utilizados, e à semelhança do referido na bibliografia (47, 56, 57), contemplou-se a introdução de degenerações nas posições onde se verificou maior variabilidade entre as sequências comparadas (posições polimórficas). Esta medida teve como objectivo favorecer a amplificação quer a nível intra-genotípico, quando considerados os *primers* específicos, quer inter-genotípico, quando considerados os *primers* de amplificação universal. A juntar a isto, e como descrito nos capítulos anteriores, no desenho dos *primers* seguiram-se ainda algumas recomendações gerais como sejam, entre outras, tamanho (entre 22-25 nt), conteúdo C/G (30 a 60%) e temperatura de desnaturação ($\geq 65^{\circ}\text{C}$). No entanto, foi frequente a necessidade de introduzir pequenas excepções a estas regras gerais. Estas resultaram da dificuldade em cumprir com todas as disposições previamente estabelecidas, tendo sido fortemente condicionadas pela

natureza nucleotídica (polimorfismos) da região em estudo. Perante os resultados obtidos, algumas destas exceções serão particularmente exploradas ao longo deste capítulo, sendo sugeridas alterações, ou mesmo discutidas algumas das alternativas avaliadas experimentalmente.

Devido à complexidade do trabalho em questão, optou-se por testar a funcionalidade dos *primers* de amplificação universal (para amplificação independente da natureza genética da sequência viral) e específicos (para amplificação específica de sequências virais, neste caso para os subtipos B, G e CRF02_AG) recorrendo, inicialmente, a uma análise genérica. Esta envolveu a utilização de estirpes virais de referência (sob a forma de clones plasmídicos) e incluiu a visualização, em gel de agarose, dos produtos amplificados por PCR. Num passo preliminar procurou-se avaliar a integridade do DNA correspondente às amostras de referência. Nesta fase, optou-se por amplificar vários fragmentos de DNA compreendendo sequências distribuídas ao longo do genoma viral, utilizando conjuntos de *primers* de amplificação testados, com sucesso, em trabalhos anteriores (29, 30, 78, 104). Estes permitiram amplificar regiões distintas ao longo do genoma viral, cobrindo regiões nos extremos 5'(Gag) e 3'(Nef), bem como em regiões centrais do genoma viral (RT, Env). As amplificações das sequências virais foram conduzidas segundo um protocolo experimental no qual a utilização de uma temperatura de hibridação de 50°C teve como objectivo maximizar as hipóteses de amplificação pretendidas (1, 50). Atendendo aos tamanhos dos fragmentos esperados, os resultados da amplificação foram visualizados após electroforese em gel de agarose convencional (44). Esta etapa revestiu-se de particular importância na medida em que permitiu avaliar o sucesso dos *primers* universais desenhados exclusivamente para este trabalho, e a sua potencial aplicabilidade nas amplificações planeadas para etapas posteriores.

A etapa seguinte pressupôs a utilização do mesmo conjunto de amostras de referência, mas desta feita visando a amplificação de sequências virais utilizando, para tal, os *primers* de amplificação universal externos desenhados neste trabalho. De uma forma geral, observou-se que, para todas as regiões alvo consideradas, os *primers* universais externos originaram produtos de amplificação detectáveis em gel de agarose para as várias regiões em estudo. Para além de nos ter permitido constatar a eficácia dos *primers* de amplificação em causa, esta etapa forneceu-nos informação inestimável

sobre a forma como as reacções de amplificação eram condicionadas pela quantidade da matriz e/ou temperaturas de hibridação utilizadas. A escolha de temperaturas de hibridação superiores às utilizadas nos ensaios preliminares teve como objectivo salvaguardar a possibilidade de hibridações inespecíficas, face à extensa dimensão dos fragmentos utilizados como DNA matriz (para a primeira reacção de amplificação cerca de 10000pb, correspondendo ao genoma completo do HIV-1). Ainda assim, para cada amostra e por cada região, as amplificações foram realizadas em triplicado, correspondendo estas a 3 diferentes temperaturas de hibridação (55°C, 58°C e 60°C). Esta estratégia teve como objectivo conhecer os limites do sistema, tendo em conta que o ensaio seria posteriormente aplicado a um conjunto de amostras clínicas. Contudo, nesta fase não se detectaram alterações no padrão de amplificação em função das três temperaturas utilizadas.

O desenho dos *primers* de amplificação universal teve por base o reconhecimento no alinhamento múltiplo de sequências, que para tal foi construído, de regiões o mais conservadas possível quando considerados todos os subtipos/CRF de HIV-1 nele compreendidas. Para o efeito, permitiu-se a inclusão de nucleótidos contendo bases nucleotídicas alternativas (posições degeneradas) nas sequências dos *primers* cobrindo, assim, eventuais variações inter-genotípicas. Os *primers* deveriam apresentar um tamanho suficientemente grande para minimizar uma potencial hibridação inespecífica com outras regiões do genoma viral que não a região alvo pretendida (1, 100). No entanto, alguns destes (ver anexo 2) ostentavam um tamanho superior ao normalmente recomendado (1, 50, 115) e temperaturas de desnaturação inferiores a 65°C. Para além disto evitou-se, sempre que possível, permitir a presença de mais do que 3 nucleótidos de citidina ou guanosina consecutivos nas extremidades 3' dos *primers*, e a inclusão de sequências com várias repetições consecutivas das mesmas bases nucleotídicas (máximo 4) (1). Para o sucesso das reacções de PCR é aconselhada a utilização de *primers* que apresentem temperaturas de desnaturação $\leq 65^{\circ}\text{C}$, onde o intervalo de temperaturas ideal se situa entre os 52-58°C. Por seu lado a presença de pelo menos 3 bases nucleotídicas (C/G) nas extremidades habitualmente confere uma maior especificidade devido ao número de pontes de hidrogénio estabelecidas, contudo a presença de um maior número de bases deve ser evitada (92, 97). À semelhança do que foi anteriormente descrito para os *primers* universais externos, as amplificações em

que foram utilizados os *primers* universais internos foram conduzidas a várias temperaturas de hibridação (55°C, 58°C e 60°C) tendo sido verificada uma ligeira diminuição na quantidade do produto amplificado (visível em gel de agarose) no limite superior de temperatura utilizado. Esta situação embora esteja dependente da temperatura de desnaturação dos *primers* está directamente relacionada com o facto de que temperaturas de hibridação elevadas conduziram a um menor número de hibridações (*primer*/DNA alvo) resultando numa diminuição do produto amplificado. Para casos onde a temperatura de desnaturação dos *primers* seja extremamente elevada (superior a 65°C), a quantidade de energia necessária, neste caso fornecida pela temperatura poderá ser incomportável dentro do sistema utilizado. Nesta situação hibridações inespecíficas serão favorecidas, ressalvando ainda o facto que a temperatura do sistema será ideal para a maioria das enzimas utilizadas para a polimerização (97). Por outro lado, a utilização de temperaturas de hibridação inferiores originam o efeito contrário, podendo mesmo desencadear hibridações inespecíficas (50, 92). Outro ponto não menos importante para o sucesso destas amplificações, e de acordo com o que foi descrito nos capítulos anteriores, as reacções de amplificação dos fragmentos internos foram efectuadas a partir de 1µl do produto obtido na reacção antecedente (fragmentos internos) diluído na razão de 1:1000. Este passo resultou do facto de se ter verificado empiricamente que a elaboração das reacções directamente a partir de 1µl do produto amplificado originava, de forma reprodutível, amplificações inespecíficas ou ausência de fragmentos visíveis em gel de agarose. Neste caso, um excesso de DNA matriz terá impedido a correcta amplificação exponencial dos fragmentos desejados (28, 90). Por outro lado, para as reacções em que não foram detectadas quantidades de produto de amplificação (fragmentos externos) visíveis em gel de agarose, optou-se por iniciar as segundas reacções de amplificação universal a partir de 5µl do produto amplificado na primeira reacção.

Estes dados permitiram constatar a possibilidade de amplificação das diferentes regiões alvo do genoma do HIV-1 mediante a utilização de ambos os tipos de *primers* universais como iniciadores de PCR. No seguimento do procedimento experimental idealizado, e perante os resultados obtidos foram, então, testadas condições de amplificação selectivas (também designadas de amplificações específicas). Estas eram dependentes, agora, da identidade genética (subtipo/CRF) de cada uma das regiões do

genoma viral em análise. Assim sendo, após a identificação de regiões conservadas para o mesmo subtipo e distintas inter-genotipicamente, seleccionou-se, de entre as possíveis, um conjunto para as quais as sequências virais serviriam de alvo para os *primers* específicos. Tendo em vista as amplificações pretendidas, e contrariamente à apresentação convencional de um ensaio de MHA, recorreu-se à utilização combinada de um *primer* específico e um outro universal. Este resultou na optimização de condições experimentais onde todas as combinações (amostra/região) foram efectuadas num formato de 6 reacções individuais, nas quais se variou a temperatura de hibridação (55, 58, 60, 63, 66 e 69°C). Através da utilização deste gradiente pretendeu-se seleccionar a(s) temperatura(s) de hibridação ideal(is) para a realização dos ensaios de amplificação específicos. Embora o sucesso deste tipo de ensaios não se deva, exclusivamente, às temperaturas de hibridação e *primers* utilizados, sendo também importantes as concentrações respectivas dos reagentes, para o delineamento dos *primers* específicos, e sempre que possível, seguiram-se um conjunto de recomendações com vista a evitar potenciais amplificações inespecíficas. Para além dos parâmetros já tidos em conta na definição dos *primers* universais (presença de posições degeneradas, conteúdo C/G, temperaturas de desnaturação, tamanho, ausência de repetições consecutivas das mesmas bases e mais do que 3 bases C ou G na extremidade 3' dos *primers*), foram consideradas as sequências que apresentassem no mínimo mais do que três nucleótidos diferentes entre os respectivos subtipos/CRF. No entanto, para alguns dos *primers* de amplificação específica (ver anexo 2), não foi possível contemplar exactamente todos estes parâmetros. Por exemplo, de modo a que os *primers* específicos incluíssem um número mínimo de diferenças (pelo menos 3 nucleotídeos distintos quando comparadas as sequências de estirpes virais de subtipos/CRF distintos, tal como definido na literatura (47, 56, 57)), e ao mesmo tempo permitissem a inclusão de posições degeneradas (cobrindo eventuais diferenças na sequências virais para um mesmo subtipo/CRF), permitiu-se um tamanho superior ao desejado e/ou um conteúdo C/G e/ou temperaturas de desnaturação superiores aos limites previstos.

Para a maioria das regiões (Gag, Pr, RT, Rev e Gp41), e para todas as amostras de referência utilizadas, foi possível constatar a amplificação específica de um fragmento de dimensões aproximadas ao esperado a, pelo menos, uma das temperaturas de hibridação utilizadas. Em particular para a região Gag, e contrariamente ao

verificado para as restantes regiões do genoma viral, apenas foi possível obter amplificações específicas no limite superior das temperaturas de hibridação utilizadas (69°C). O tamanho (28nt) e a própria constituição nucleotídica dos *primers* utilizados, que no seu conjunto apresentaram elevadas temperaturas de desnaturação, espelha a dificuldade no seu delineamento, o que, em parte, resultou do facto desta região apresentar-se como uma das mais conservadas do genoma viral (54, 116).

Para a região que codifica a integrase, numa primeira abordagem assistiu-se a uma dificuldade em identificar, correctamente, os produtos amplificados para as amostras de referência, independentemente das temperaturas de hibridação utilizadas. Esta situação foi devida às pequenas dimensões dos amplicões gerados, ainda que se tenha optado, excepcionalmente, por resolver os mesmos em gel de poliacrilamida. Ainda que este tipo de gel permita separar fragmentos de menores dimensões comparativamente aos geis de agarose (16), a análise dos resultados obtidos, quando combinada com as pistas referentes aos controlos negativos de amplificação, permitiu a visualização de fragmentos de pequenas dimensões (entre os 25 e 35pb) em todas as pistas, sugerindo a possibilidade de ocorrência de dímeros de *primers*, e impossibilitando uma clara discriminação a partir desta análise. De forma a clarificar esta situação, redesenhou-se um novo *primer* universal interno de forma a permitir a obtenção de um amplicão de maiores dimensões. No entanto, os produtos obtidos vieram a revelar a presença de um padrão de amplificação inespecífico. Tal situação foi traduzida pelo facto de que o *primer* específico para sequências virais do subtipo B permitir a sua hibridação heteróloga com sequências de referência do subtipo G e da forma recombinante CRF02_AG à temperatura máxima de hibridação utilizada (69°C). Depois de reanalisada, em detalhe, a estrutura do *primer* específico para o subtipo B, verificou-se que apesar deste apresentar apenas 25nt de tamanho, a sua temperatura de desnaturação era muito superior ao máximo recomendado (65°C) o que poderá estar na origem dos resultados obtidos (50, 92, 100). De modo a ultrapassar esta situação, optou-se por se redesenhar um novo conjunto de *primers* específicos, para os quais foi dada particular atenção às situações apontadas anteriormente (tamanho do fragmento esperado, redução da temperatura de desnaturação) juntamente com a inclusão, de forma propositada, de nucleótidos não homólogos na sequência do *primer* específico para as sequências do subtipo G e forma recombinante CRF02_AG, tal como sugerido

em estudos semelhantes (22). A análise dos produtos amplificados obtidos mediante a utilização deste segundo grupo de *primers* revelou que, apesar do planeamento efectuado, o *primer* específico para o subtipo B permitia a amplificação de fragmentos heterólogos (subtipo G e CRF02_AG), para além do que o *primer* específico para as sequências do subtipo G originava fragmentos de tamanhos superiores aos esperados quando a matriz de amplificação utilizada correspondia a DNA do clone molecular correspondendo à forma CRF02_AG. Estes resultados revelaram-se reprodutíveis apesar da grande amplitude de temperaturas de hibridação utilizadas (de 55 a 69°C). No caso do *primer* de amplificação específica para o subtipo B os resultados obtidos podem ter sido originados pelo facto do maior número de diferenças estar representado na extremidade oposta à extremidade na qual a enzima (*Taq*) se liga, permitindo, assim, que uma hibridação inespecífica apenas desta extremidade (3' do *primer*) inicie a polimerização (97, 115). Por seu lado, na origem dos fragmentos de tamanho superior ao desejado, obtidos com o *primer* específico para o subtipo G, a reduzida dimensão do próprio *primer* utilizado (apenas com 19nt), poderá ter promovido a ocorrência de hibridações inespecíficas (27, 50, 100). As dificuldades encontradas no delineamento dos *primers* em questão, e os sucessivos resultados inadequados, reproduzem o facto desta região apresentar-se, tal como a região que codifica Gag, como uma das mais conservadas do genoma (11) Por estas razões, e ainda por limitações impostas pelo tempo disponível para a realização deste trabalho, optou-se por não prosseguir com a análise desta região do genoma do HIV-1.

De forma semelhante ao que se veio a verificar com a região que codifica a glicoproteína Gp120, o conjunto de *primers* desenhados inicialmente (sufixo 1) originou, de uma forma geral, produtos de amplificação independentemente da identidade genética da matriz de referência utilizada, isto é, tanto foram obtidos produtos de amplificação a partir da amostra de referência homólogas (ex: *primer* B/referência de subtipo B) como heterólogas (ex: *primer* G/referência de subtipo B) relativamente ao *primer* específico utilizado. No entanto, uma vez mais, as reduzidas dimensões dos fragmentos amplificados (<50pb) dificultaram a correcta visualização de possíveis amplicões específicos. Para além disto, e com a experiência adquirida durante a análise da região que codifica a integrase viral, verificou-se que todos os *primers* de amplificação específica para a região codificante de Gp120 apresentavam temperaturas

de desnaturação muito superiores aos desejado o que, independentemente do tamanho dos fragmentos, daria origem, possivelmente, a dificuldades na interpretação dos resultados do ensaio. Face a esta situação foram redesenhados 6 novos *primers* específicos (dois conjuntos: Gp120-sufixo 2 e 3, ver anexo 2) juntamente com dois novos *primers* de amplificação universais. Os novos *primers* de amplificação específica foram delineados de forma a hibridar com regiões distintas das definidas anteriormente, tendo sido orientados no sentido reverso relativamente aos primeiros (e à sequência codificante). Tal facto foi considerado uma vez que a maioria das diferenças entre cada um dos *primers* e as sequências de vírus pertencentes a subtipos não homólogos situavam-se na extremidade 3'. Estes conjuntos diferiam, essencialmente, no número de nucleótidos, onde a sequência dos *primers* Gp120-sufixo 2 compreendia a sequência dos *primers* Gp120-sufixo 3. À semelhança do anteriormente descrito, neste último conjunto foi, propositadamente, contemplada a inclusão de nucleótidos não homólogos (i.e. *à priori* não complementares à sequência alvo) de modo a desfavorecer hibridações com as sequências virais de vírus de um determinado subtipo, relativamente a outros (22).

A análise dos fragmentos amplificados com o conjunto Gp120-sufixo 2 revelou a presença de amplicões específicos quando o *primer* específico para sequências do subtipo B foi utilizado em conjunto com uma matriz do mesmo subtipo, a temperaturas de hibridação superiores a 63°C. Por outro lado, o *primer* específico para sequências do subtipo G originou amplicões quando utilizadas como matriz amostras de referência do subtipo G, ou da forma CRF02_AG. No entanto, para as reacções de amplificação em que foram utilizadas temperaturas de hibridação superiores a 60°C, verificou-se uma diminuição visível (em gel de agarose) da quantidade do produto amplificado quando a matriz utilizada e *primer* específico eram não-homólogos (ex: *primer* G/referência de subtipo B ou CRF02_AG). De uma forma semelhante, o *primer* específico para sequências da forma recombinante CRF02_AG permitiu a obtenção de amplicões às temperaturas de 55 e 60°C, independentemente das referências utilizadas corresponderem a sequências dos subtipos B ou G. Para temperaturas de hibridação iguais ou superiores a 63°C não se visualizavam produtos de amplificação para as referências do subtipo B, contrariamente ao resultado observado quando utilizada a referência de subtipo G, para a qual se registaram amplificações heterólogas até à

temperatura de hibridação máxima ensaiada ao longo deste trabalho (69°C). O facto de que os produtos das amplificações específicas terem sido apenas visualizados a temperaturas de hibridação acima dos 63°C (excepto para combinação *primer* específico para CRF02_AG/amostra de referência do subtipo G) podem dever-se ao facto dos *primers* utilizados apresentarem temperaturas de hibridação próximas do limite estipulado (65°C).

Ainda para esta mesma região, e para o terceiro conjunto de *primers* específicos (Gp120-sufixo **3**), a análise dos produtos de amplificação revelou a presença de um amplicão para cada um dos *primers* específicos (subtipo B e CRF02_AG) e respectivas amostras homóloga (à temperatura de hibridação de 63°C). Contudo, simultaneamente, foi detectada a presença de fragmentos inespecíficos de dimensões superiores ao esperado em alíquotas de todas as reacções de amplificação, incluindo as que foram realizadas à temperatura máxima utilizada ao longo destes ensaios (69°C). Na origem dos fragmentos de dimensões superiores ao desejado estão hibridações inespecíficas que podem, muito provavelmente, ser consequência das pequenas dimensões dos *primers* utilizados (20nt) e/ou da enorme variabilidade genética que caracteriza a maior parte da sequência codificante de Env (52). As reacções de PCR realizadas com o *primer* específico para o subtipo G traduziram-se na detecção de amplicões quando utilizadas referências dos subtipo G ou da forma CRF02_AG, apesar da grande amplitude térmica utilizada. Esta situação pode ter resultado da presença de 3 nucleótidos de citidina consecutivos na extremidade 3' do *primers* ou, simplesmente, ao facto da sequências complementares às dos *primers* específicos se localizarem numa zona próxima do ponto de recombinação (subtipo G/subtipo A) que caracteriza, nesta zona do genoma viral, a CRF02_AG.

De um modo geral verificou-se que *primers* que apresentassem temperaturas de desnaturação elevadas ($\geq 65^{\circ}\text{C}$) ou dimensões inferiores a 22nt geralmente originavam amplificações inespecíficas até à temperatura máxima de hibridação de 69°C utilizada. Por outro lado, reacções de amplificações realizadas a elevadas temperaturas (69°C) devem ser evitadas uma vez que estas se situam muito perto da temperatura a que a polimerase termorresistente utilizada (115) tem actividade máxima. Para além disto, e das várias alterações efectuadas com vista a melhorar a especificidade das reacções de amplificação, realça-se a extrema dificuldade em obter conjuntos de *primers* para esta

região (Gp120) do genoma viral. Da análise do alinhamento múltiplo de sequências construído foi possível constatar a escassez de sequências nucleotídicas que reunissem as características necessárias para o delineamento de *primers* específicos. De forma a cobrir toda a variabilidade intra-genotípica seriam necessários *primers* excessivamente representados por posições degeneradas, o que, no limite, conduziria, muito provavelmente, a hibridações não desejadas. Todas estas limitações suportam o facto de esta região ser caracterizada como aquela que apresenta maior variabilidade nas sequências virais, tal como amplamente descrito na literatura (52, 116).

A utilização de uma abordagem experimental pressupondo a amplificação/identificação de sequências virais em três passos sucessivos pode ser entendida como desvantajosa, quando comparada com os ensaios convencionais de MHA. Em termos práticos traduz-se num manuseamento extra dos produtos obtidos, o que, por si só, aumenta as probabilidades de contaminações, para além do consumo de tempo associado. No entanto, e por outro lado, permite a amplificação de sequências virais de amostras com baixa carga viral e simultaneamente, restringe o tamanho dos fragmentos que serviriam de DNA matricial para a reacção de PCR de carácter específico. Por outras palavras, para uma primeira reacção é utilizado como DNA matricial o genoma total do HIV-1, do qual são apenas amplificadas regiões parciais (fragmentos universais externos), as quais serão utilizadas como DNA molde para uma segunda reacção, ainda de carácter universal (fragmentos universais internos). Uma vez obtidos os segundos fragmentos, estes servirão de alvo para os *primers* de amplificação específica. Uma vez que se partiu de um fragmento inicial de DNA com cerca de 10000pb, para um fragmento de menores dimensões (cerca de 200pb) para alvo dos *primers* específicos, todo este procedimento restringe, até certo ponto, as probabilidades de hibridações inespecíficas, aumentando a especificidade dos resultados obtidos.

A funcionalidade dos *primers* específicos foi ainda avaliada através de um ensaio de competição de matrizes. Basicamente, nestes procedeu-se à amplificação de diferentes regiões do genoma viral utilizando amostras de referência e uma abordagem experimental baseada na PCR em tempo-real. Em termos práticos utilizou-se informação obtida nos ensaios de PCR anteriormente descritos, nomeadamente no que se refere à definição das temperaturas de óptimas de hibridação, para as quais haviam sido obtidos amplicões específicos para as amostras de referência com os respectivos

primers homólogos. A análise das curvas de amplificação obtidas demonstrou que, para as regiões estudadas (Gag, Pr, RT, Rev, Gp120 e Gp41), apesar da presença de um competidor (DNA não homólogo) em excesso relativamente à matriz homóloga, os *primers* de amplificação específica permitiram a correcta detecção da matriz homóloga. A inclusão de um controlo positivo (presença exclusiva de DNA alvo homólogo relativamente ao *primer* específico utilizado) permitiu avaliar eventuais influências do competidor. As observações efectuadas corroboram os resultados de amplificação específica obtidos anteriormente.

A opção tomada relativamente à utilização, neste tipo de ensaios, de um “representante” de cada subtipo/CRF baseou-se no facto de que as amplificações específicas, descritas na secção anterior, terem revelado resultados semelhantes para todas as amostras de um mesmo subtipo/CRF utilizadas. Contrariamente aos ensaios efectuados por PCR convencional (com consequente visualização de produtos de amplificação em gel de agarose), as quantidades relativas das diferentes matrizes utilizadas foram equilibradas de modo a que se apresentassem numa concentração semelhante por cada região analisada. Embora, no seu fundamento, este tipo de ensaios deva permitir a identificação da natureza genética da sequência viral dentro de uma gama de concentrações/número de cópias de genoma viral, de facto, um aumento exponencial de sinal fluorescente a partir de um ponto Ct é ainda função da concentração de DNA alvo presente na mistura de amplificação. Para além disto, ao longo deste trabalho verificou-se que, nalgumas condições (elevadas concentrações de matriz), a presença de uma matriz não-homóloga nas misturas de reacção pode dar origem a amplificações indesejadas, contribuindo para resultados inespecíficos. Para o sucesso dos resultados obtidos é ainda importante salientar que a mistura de PCR comercial utilizada incluía a presença de uma polimerase de DNA termorresistente e quimicamente modificada, que se apresentava inactiva à temperatura ambiente (mas passível de ser activada pelo calor (*hot start polymerase*), impedindo que hibridações não específicas dos *primers* pudessem resultar em amplificações inesperadas. Esta característica garante uma maior especificidade nas reacções de amplificações de DNA pretendidas. Pelas razões referidas anteriormente, não foram efectuados ensaios de competição para a região da integrase.

Uma vez otimizados os protocolos de amplificação específica quer utilizando uma estratégia baseada em PCR convencional e análise de produtos em gel de agarose, quer baseada em PCR em tempo-real, e um conjunto de amostras de referência, decidiu-se avaliar o sucesso destes protocolos usando um conjunto de amostras clínicas já analisadas no âmbito de uma série de estudos de caracterização genética de estirpes de HIV-1 realizados, anteriormente, no nosso laboratório. Em relação à amplificação dos fragmentos universais externos, a análise em gel de agarose dos produtos obtidos revelou, em particular para as regiões de Gag e Pr, a presença dos fragmentos de tamanho esperado, e para a grande maioria das amostras biológicas, desde a primeira reacção de amplificação. Contudo, para algumas amostras, não foram detectados produtos amplificados visíveis, facto que se verificou, como regra geral, para as restantes regiões analisadas (RT, Rev, Gp120 e Gp41), num total de 50 amostras. Para estas, apenas foi possível detectar os produtos amplificados numa segunda reacção de amplificação, ainda de carácter universal, correspondendo os amplicões obtidos aos fragmentos universais internos. Ambas as reacções de carácter universal foram realizadas à temperatura de hibridação mínima ensaiada (55°C) para as amostras de referência, de modo a maximizar as hipóteses de amplificação. A obtenção de amplicões universais externos para as regiões de Gag e Pr apenas com um protocolo de PCR foram surpreendentes. De facto, na bibliografia é frequentemente documentado a utilização de estratégias de amplificação por *nested*-PCR de modo a obter quantidades suficientes de produtos amplificados a partir de amostras clínicas. Esta situação é particularmente referida quando a análise é baseada em amostras biológicas de indivíduos submetidos a terapia antirretroviral, na medida em que esta reduz, drasticamente, o número de cópias de genoma viral nos indivíduos infectados (21, 40, 87). Por esta razão, optou-se por desenvolver uma estratégia com várias reacções de amplificação, nas quais os produtos de uma serviriam de matriz para a reacção seguinte, apesar do facto de a quantidade de matriz não constituir um factor limitante quando analisadas as amostras de referência.

Assim, após terem sido amplificados os fragmentos universais internos para todas as regiões do genoma das diferentes estirpes de HIV-1 presentes nas amostras clínicas em estudo, iniciaram-se os ensaios de amplificação específica. Estes foram realizados recorrendo quer à detecção, em gel de agarose, dos produtos amplificados

PCR convencional, quer por recurso ao PCR em tempo-real. Uma primeira análise dos resultados revelou padrões de amplificação inespecífica para todas as regiões estudadas (excepto a região de Gag) para a maioria das amostras. Perante esta situação, e uma vez que estas reacções tinham sido realizadas às temperaturas de hibridação definidas para as amostras de referência, optou-se por uma re-optimização dos ensaios através de um incremento gradual da temperatura de hibridação. Esta medida mostrou-se útil uma vez que, para a maioria das amostras, permitiu a obtenção de um resultado de amplificação específico com apenas um dos *primers* de amplificação disponíveis. Em particular para a região de Gag, por ter sido a região que inicialmente apresentou o maior número de amplificações exclusivas apenas para um dos *primers*, e pelo facto de estarmos a trabalhar no limite máximo de temperaturas de hibridação definido (69°C), não se fizeram alterações ao protocolo inicial. No entanto, seria importante rever, detalhadamente, a própria composição dos *primers* específicos, juntamente com o alinhamento de sequências construído, tendo em vista delinear (caso isto seja possível) novos *primers*, reduzindo o tamanho dos mesmos e, conseqüentemente, as respectivas temperaturas de desnaturação.

De facto, o padrão de amplificações registado foi, até certo ponto, previsível, dado terem sido analisadas amostras clínicas. Nestas é expectável que seja grande a variabilidade das sequências das estirpes de HIV-1 nelas presentes, variabilidade esta, que deverá contribuir para o resultado obtido (4, 116). Por outro lado, a utilização de um maior número de amostras de referência deveria ter permitido, à partida, explorar este tipo de situação. Quando comparados os resultados anteriores (PCR convencional/visualização em gel de agarose) com a análise das curvas de amplificação específica por PCR em tempo-real observou-se uma redução considerável do número de amostras para as quais foram obtidas, nesta última estratégia, amplificações com mais do que um *primer* específico. Salientado ainda que os mesmos foram uma constante para todas as regiões tendo em conta o mesmo conjunto de amostras clínicas.

Estes resultados podem ser explicados pelo facto de que em gel de agarose são visualizados os produtos finais de um processo de amplificação, enquanto que a PCR em tempo-real nos permite ter uma ideia, ao longo dos vários ciclos de amplificação, da forma como esta evolui ao longo do tempo. Por outras palavras, contrariamente ao registo, em gel de agarose, de uma amplificação dupla/tripla, as reacções de PCR em

tempo-real permitiram a observação de um padrão de amplificação discriminatório, definido por uma curva de amplificação exponencial com um dos *primers* específicos. A observação, em simultâneo, de curvas de amplificação como resultado da utilização de combinações não-homólogas entre *primers* e matrizes foi, na maioria das vezes, restrita ao final do protocolo de amplificação. Para um conjunto reduzido de amostras (ver anexo 1) e em particular para algumas das regiões (maioritariamente para a região de Gp41), não foram detectados produtos amplificados pela estratégia por PCR convencional independentemente do *primer* específico utilizado. Esta situação pode dever-se ao facto de que esta estratégia não permite visualizar quantidades reduzidas de produto amplificado, ao contrário da abordagem por PCR em tempo-real permitiu detectar algumas destas nas condições de reacção utilizadas (5, 41).

Apesar disto, algumas das amostras revelaram resultados de amplificação indiscriminados, independente da estratégia utilizada. Em particular, a região de Gp120 foi aquela para a qual se verificou um menor número de amplificações específicas (68% por PCR convencional e 76% por PCR em tempo-real) tendo em consideração apenas o segundo conjunto de *primers*. Embora o terceiro conjunto tenha originado um número superior de resultado exclusivos para apenas um dos *primers*, em relação ao segundo conjunto por PCR/gel agarose convencional, pelas razões apresentadas anteriormente (presença de fragmentos de dimensão superior ao desejado, amplificações heterólogas para o *primer* específico do subtipo G quando foram quando utilizados com as amostras da CRF02_AG.), o mesmo não foi avaliado numa estratégia por PCR em tempo-real. Na base desta opção esteve o facto de que o sinal de fluorescência emitido na presença *SYBR*[®] *Green* I seria influenciado pela amplificação de fragmentos de dimensões superiores ao desejado, originando falsos positivos (61, 115). Mais uma vez os resultados obtidos muito provavelmente vêm reflectir a grande variabilidade que caracteriza esta região do genoma.

Paralelamente, foi ainda experimentada uma abordagem a dois passos, em que para a última reacção de carácter específico foram utilizados os produtos obtidos para os fragmentos externos. Para tal recorreu-se ao PCR em tempo-real e a um conjunto aleatório de amostras clínicas. A análise dos resultados demonstrou a amplificação de fragmentos específicos para as regiões de Gag, Pr, e RT. Contudo, para as restantes regiões (Rev, Gp120 e Gp41) observaram-se, curiosamente, curvas de amplificação

inespecíficas para todos os ensaios realizados. Estes resultados podem, em parte, ser explicados pelo facto da extremidade 5' do genoma incluir as regiões mais conservadas, contrastando com as restantes regiões (extremidade 3'), caracterizadas por uma maior variabilidade das sequências nucleotídicas (52). Esta última característica pode ser responsável por um conjunto de hibridações inespecíficas que contribuíram para detecção de um sinal inesperado de amplificação. Por outro lado, a maioria das reacções para a extremidade 5' do genoma do HIV, com excepção da região RT, foram iniciadas com 1µl de um produto obtido na primeira reacção de amplificação depois de diluído na razão de 1:1000. Para as restantes regiões, e por não terem sido detectados produtos em gel de agarose, optou-se por incluir nas misturas de reacção 5µl dos produtos amplificados (fragmentos externos). A utilização de um maior volume (cerca de ¼ do total do produto amplificado na reacção antecedente) poderá ser responsável pelo transporte de *primers* não utilizados na reacção anterior, cuja presença, nas reacções subsequentes poderá ter contribuído para os resultados obtidos.

A título experimental foi ainda realizado um conjunto de amplificações por PCR em tempo-real, mas desta feita recorrendo a sondas de hidrólise, seguindo a estratégia que, de facto, caracteriza os ensaios de MHA descritos na bibliografia (46, 56, 57). Pelos custos associados à síntese das sondas em causa, e pelas limitações temporais impostas à realização deste trabalho, estas reacções foram dirigidas apenas para a região de Gp41 e para um conjunto de amostras, aleatoriamente seleccionadas. A escolha desta região deveu-se ao facto de ser, por entre as várias regiões do genoma do HIV-1 analisadas, aquela para a qual maior número de resultados de amplificação reprodutíveis por entre as abordagens efectuadas. A análise das curvas de amplificação revelou que para 14 (52%) das amostras em análise foi possível detectar um sinal de hibridação específico para um dos subtipos virais (B ou G), sendo que todos os resultados coincidiram com os obtidos a partir das estratégias citadas anteriormente. Para as restantes amostras, inesperadamente, não foram registados quaisquer sinais indicadores de amplificação de sequências virais, independentemente da sonda utilizada. Estas observações podem resultar de uma própria limitação do método, uma vez que as condições utilizadas foram optimizadas para um procedimento (PCR convencional e visualização em gel de agarose e/ou PCR em tempo-real) com parâmetros distintos do contexto actual. Por outras palavras, as estratégias anteriores tinham como objectivo a

amplificação de segmentos de genoma viral com *primers* específicos, ao passo que nesta abordagem as sondas vão hibridar com a sequência homóloga, e mais tarde sofrer hidrólise, ocorrendo a emissão de fluorescência aquando da polimerização, iniciada pela presença na mistura de reacção de um outro par de *primers* (neste caso foram utilizados os *primers* universais internos). Em termos temporais, a hibridação dos *primers* internos deverá ter lugar posteriormente à hibridação da sonda, de modo que, ao iniciar-se a polimerização, a sonda esteja convenientemente ligada à sua sequência alvo. Para tal as sondas deverão apresentar temperaturas de desnaturação superiores às dos *primers* de amplificação (47, 56, 57). As sondas específicas para a realização deste ensaio foram desenhadas de forma a apresentarem como requisito mínimo, a mesma sequência nucleotídica dos *primers* específicos. Na bibliografia alguns estudos referem que para um desempenho óptimo em função do delineamento dos *primers* e sondas, estes ensaios são realizados normalmente à temperatura de hibridação de 55°C (56, 57), contudo, neste caso a temperatura de hibridação utilizada foi de 68°C, uma vez, que esta temperatura permitiu a amplificação discriminatória de sequências virais das amostras analisadas pelas estratégias anteriores. De facto a realização do ensaio a 68°C deveria, em condições normais, evitar a hibridação dos *primers* de amplificação universal. Não obstante, tal não se verificou pelo facto dos mesmos apresentarem, excepcionalmente temperaturas de desnaturação elevadas. Esta situação resultou, não da sua dimensão, mas do facto desta região ser particularmente rica em nucleótidos de C/G. Paralelamente, foram ainda efectuados ensaios de tipo *multiplex* nos quais ambas as sondas de hidrólise foram utilizadas, simultaneamente, na mesma reacção de amplificação. O facto de as sondas estarem marcadas propositadamente com fluoróforos diferentes, JOE e ROX, para as estirpes virais dos subtipos B e G respectivamente (leitura efectuada conjuntamente em canais diferentes), permitiu a realização deste ensaio. Para além disto foram escolhidas duas das amostras para as quais foram obtidos resultados específicos distintos nas abordagens anteriores. Os resultados obtidos neste tipo de ensaio demonstraram a detecção de sinal específico para cada uma das amostras em análise tendo, para ambas, sido registado o mesmo padrão revelado nos ensaios realizados anteriormente. Embora esta abordagem não esteja descrita na bibliografia e tenha sido executada sem qualquer tipo de avaliação quanto à sua reprodutibilidade, os resultados obtidos demonstram a potencialidade deste tipo de estratégia, e a

possibilidade de utilização de várias sondas em simultâneo numa mesma mistura de reacção.

A optimização das reacções específicas ao longo deste trabalho foi efectuada essencialmente através do incremento das temperaturas do hibridação face ao TM dos próprios primers, ou ainda pelo redesenho dos primers. Esta situação resulta de vários factores, entre os quais, o objectivo claro que desenvolver um método robusto, que permita dentro de certos parâmetros o rastreio de um conjunto de amostras clínicas. Por outro lado, o recurso a misturas de reacção prontas a usar, embora ofereça algumas vantagens, uma vez que são mais fáceis de preparar e reduzem a probabilidade de contaminações, por outro limitam a capacidade do utilizador interferir com as concentrações dos próprios reagentes. Como exemplo disto a presença de $MgCl_2$ é um dos constituintes essenciais das misturas, onde a sua concentração ideal traduz-se em resultados específicos de amplificação (115).

De forma a avaliar a sensibilidade e especificidade dos ensaios é necessário comparar os resultados obtidos com o método considerado como o mais robusto, a sequenciação do genoma completo. A maioria dos trabalhos publicados referindo o desenvolvimento de ensaios deste tipo, é implementada em parceria com grandes centros de investigação com enormes recursos, em que são levados a cabo estudos de epidemiologia molecular do HIV-1 (46, 47, 56, 57, 119), em particular no que diz respeito à possibilidade de sequenciação de genomas virais. Para a realização deste trabalho, não foi possível contar *à priori* com esta informação. Face a esta situação, optou-se por sequenciar vários fragmentos que cobriam as regiões analisadas anteriormente para todas as amostras clínicas em estudo. Para uma análise filogenética consistente no caso do HIV-1 é necessário que os fragmentos apresentem um tamanho mínimo (\geq a 300nt), para que a informação contida nas mesmas seja suficiente para inferir sobre a origem genética da sequência viral (48). Devido às pequenas dimensões dos fragmentos obtidos nos ensaios de amplificação específica (PCR convencional/visualização em gel de agarose ou PCR em tempo-real) foram efectuadas novas reacções de amplificação utilizando uma estratégia de *heminested*-PCR, na qual foram utilizados em conjunto um *primer* de amplificação universal externa e outro de amplificação universal interna. Os fragmentos obtidos tinham dimensões (dependendo da região) compreendidas entre os 400pb e os 680pb. Pelas limitações de tempo

impostas, e complexidade associada à realização deste trabalho, até à data da finalização deste manuscrito não foi possível apresentar os resultados da análise filogenética para todas as regiões (Gag, Pr, RT, Rev, Gp120 e Gp41) das 50 amostras em estudo. Encontram-se deste modo apresentados os resultados disponíveis até esse período, sendo que a sequenciação e análise filogenética continuam em curso. Alguns resultados são comparados ainda com os dados disponíveis resultantes de trabalhos anteriores em que se procedeu à análise de sequências virais amplificadas das mesmas amostras biológicas (HMA, análise filogenética (29, 30, 78). É importante destacar que as análises parciais efectuadas anteriormente, embora tenham origem na mesma amostra biológica, na maioria dos casos as regiões analisadas e o material biológico de origem foi distinto (RNA viral vs. DNA proviral). Como tal, foi utilizada apenas como ultimo recurso até estar disponível a informação relativamente à análise filogenética e de modo a ter uma ideia preliminar dos resultados definitivos.

A análise das diferentes árvores filogenéticas auferidas revelou que as sequências obtidas neste trabalho surgiram agrupadas com as referências dos subtipos B e G, e das formas genéticas CRF02_AG e CRF14_BG. Foi disto excepção a amostra 317 que surgiu agrupada com referências do subtipo D para a região de Gag, e da amostra 658 que surgiu agrupada com referências sub-subtipo F1 para a região Pr. Esta observação, apesar de resultar de um conjunto reduzido de amostras vem de algum modo de encontro ao descrito na bibliografia reflectindo a presença das estirpes dominantes na epidemia Portuguesa (30, 75, 76, 78).

Como referido ao longo do capítulo anterior, a maioria dos resultados obtidos através das estratégias desenvolvidas ao longo deste trabalho foram coincidentes com os dados obtidos através da análise filogenética das sequências virais amplificadas. Para um número reduzido de amostras/região, registaram-se algumas excepções, nas quais se destacam algumas divergências entre os resultados obtidos mediante utilização de diferentes estratégias experimentais. Noutros casos, a caracterização de sequências virais foi, *à priori*, dificultada pela sua projecção evidente como nós terminais de ramos longos relativamente a todas as restantes sequências nucleotídicas incluídas no mesmo grupo monofilético das árvores filogenéticas construídas. Para completar a sua caracterização, estas foram sistematicamente avaliadas quanto à sua natureza recombinante, por recurso à técnica de *bootscanning*.

Apesar de, na maioria dos casos, as abordagens utilizadas tenham permitido uma caracterização inequívoca das sequências virais em estudo, nalguns casos a complexidade genética do HIV-1 gerou algumas situações contraditórias. Uma das situações registadas deverá reflectir o grau de proximidade entre os “subtipos” virais B e D que, na realidade, deveriam ser considerados sub-subtipos de um único subtipo viral (62, 89). Tal é ilustrado, para a região Gag, com os resultados referentes à amostra 317. Esta agrupou com as amostras do subtipo D por análise filogenética, embora tivesse sido identificada como pertencente ao subtipo B por PCR. A análise por *bootscanning* efectuada confirmou a sua clara classificação para o subtipo D, mas falhou na identificação da sua origem genética na extremidade 3'. Pontualmente, como foi traduzido pela análise da amostra 471, sequências virais na região Gag agrupam com as amostras de referência do subtipo G (segregando com sequências de tipo CRF14_BG) apesar de terem originado resultados de amplificação duplos com primers específicos para o subtipo G e CRF02_AG. A análise de *bootscanning* não revelou quaisquer evidências óbvias de recombinação intergenotípica (embora para alguns pontos específicos da sequência nucleotídica tenha reflectido alguma divergência em relação as estirpes de referência utilizadas para o CRF14_BG). No entanto, observando os resultados obtidos ao longo do genoma e simultaneamente as informações fornecidas por ensaios realizados anteriormente, esta demonstra um padrão de amplificação para vários subtipos dependendo da região em estudo, onde, fenómenos de recombinação múltiplos e pelo facto da região alvo se situar perto dos pontos de recombinação podem justificar este resultado.

Algumas situações revelaram, inesperadamente, resultados aparentemente contraditórios, cuja clarificação mereceu especial atenção. É exemplo disto, a análise das sequências Gag amplificadas da amostra 176. Embora as sequências virais da mesma tenham originado amplificações específicas para o subtipo B, a análise filogenética da mesma identificou-a como pertencente ao subtipo G. O gráfico de *bootscanning* relativo à análise do fragmento amplificado evidenciou um fenómeno de recombinação entre o subtipo B (extremidade 5') e sequências G filogeneticamente semelhantes à sequência G que deverá ter estado na origem de forma recombinante CRF14_BG (extremidade 5'). Curiosamente, a região alvo de hibridação dos *primers*

específicos situa-se aproximadamente a 100pb do início da sequência utilizada para a análise de recombinantes, o que justifica os resultados apresentados.

Por outro lado, embora a utilização de *primers* específicos e a sua análise prévia por HMA tenham concordado na associação das sequências Gag amplificadas da amostra 508 ao subtipo G, a sua análise filogenética identificou-a como pertencente ao subtipo B. Contudo, a análise por *bootscanning* desta mesma sequência falhou na atribuição clara de um subtipo viral à região central da mesma, ainda que na sua maior parte esta deva compreender uma origem no subtipo B. Este resultado pode reflectir a proximidade da região alvo para a hibridação do *primers* específico com o ponto de recombinação originando assim uma amplificação para o subtipo G.

Uma situação semelhante foi revelada pela análise das sequências virais correspondentes à região Pr, as quais foram amplificadas especificamente com *primers* para o subtipo G a partir da amostra 658, sendo que a sua análise filogenética as associou ao sub-subtipo F1. Contudo, esta mesma sequência apresentou-se no extremo de um ramo evidentemente prolongado relativamente às restantes sequências do mesmo agrupamento monofilético. A avaliação da sua potencial natureza recombinante evidenciou dois possíveis locais de *crossing-over* que separam a secção da sequência com clara homologia com as referências do sub-subtipo F1, de uma secção não subtipável por ausência de semelhança com as sequências de referência utilizadas. Esta situação poderá justificar os resultados obtidos para as amplificações específicas, onde a amostra foi identificada para o subtipo G. Para a amostra 107, as amplificações específicas da região Pr (297pb) associada a uma prévia caracterização filogenética exclusivamente dirigida à Pr viral (S. Castro, comunicação pessoal) identificaram-na como pertencente ao subtipo G. No entanto, a análise filogenética de um fragmento de maiores dimensões (635pb) integrando parte da RT, tal como o que aqui foi analisado, identificou a estirpe viral em causa como pertencente ao subtipo B. Estes resultados, complementados com a observação de que esta sequência ser a primeira sequência a divergir em relação às demais incluídas no mesmo grupo monofilético, eram compatíveis com a natureza recombinante da mesma. De facto, o resultado da análise de *bootscanning* efectuada evidenciou um evento de recombinação intergenotípica caracterizado pela existência de uma região (extremidade 5') não subtipável por ausência de semelhança com as sequências de referência utilizadas. Por seu lado, a

restante sequência apresentou-se filogeneticamente aparentada com as amostras de referência do subtipo B. De facto, a análise filogenética exclusivamente da região da RT (S. Castro, comunicação pessoal) de sequências virais amplificadas da mesma amostra clínica identificaram-na como pertencente ao subtipo B. Considerando ainda o facto de que para a região Gag as sequências virais amplificadas desta mesma amostra se revelaram pertencentes ao subtipo G, estes resultados parecem indicar a existência clara de um ponto de recombinação na região de Pr, sendo que a região alvo para os *primers* específicos está no limite que permite a associação desta região do genoma desta estirpe viral ao subtipo G.

Algumas situações revelaram resultados contraditórios de difícil interpretação. Por exemplo, enquanto que a região RT da estirpe viral da amostra 508 foi amplificada de forma dupla com ambos os *primers* específicos para os subtipos B e G, esta foi agrupada com as amostras de referência do subtipo B através da sua análise filogenética. A inspecção da sua potencial natureza recombinante não revelou quaisquer indicações que sugerissem que eventos de recombinação tenham estado na sua origem. Esta situação pode descrever falta de especificidade dos *primers* na correcta identificação das sequências de HIV-1 presentes nesta amostra. Tal pode, provavelmente, ser consequência directa da variabilidade genética deste vírus e/ou traduzir a identificação de uma co-infecção e, assim, revelar a coexistência de uma população viral mista numa mesma amostra biológica. Situações semelhantes foram ainda registadas aquando da região Gp120, tal como evidenciado pela análise das sequências virais amplificadas das amostras 4132 e 4773. Ambas apresentando um perfil de amplificação duplo com os *primers* específicos para sequências dos subtipo G e da forma CRF02_AG, enquanto que a sua análise filogenética indicou a sua segregação com as amostras de referência pertencentes à CRF02_AG. Este resultado veio a ser confirmado através da inspecção do seu potencial recombinante por *bootscanning*. Com base nesta observação, e juntamente com a informação recolhida para outras regiões do genoma, estes resultados parecem descrever, muito provavelmente, uma situação de falta de especificidade dos *primers* específicos para o subtipo G. Contudo pelo facto de estarmos a trabalhar com vírus recombinantes existe sempre a possibilidade da região alvo de hibridação corresponder ao subtipo G.

Finalmente para um conjunto de cinco amostras (sufixo PG) os resultados das amplificações específicas por PCR convencional/visualização em gel de agarose e PCR em tempo-real (embora não tenham sido realizadas análises filogenéticas) caracterizaram-se na grande maioria dos casos por amplificações inespecíficas com todos os *primers* específicos ou simplesmente não se registaram amplificação de todo para as regiões do genoma analisadas. Estas amostras não tinham origem Portuguesa (mas sim Moçambicana), tendo sido propositadamente incluídas neste trabalho pelo facto de terem servido de material de partida para a caracterização de estirpes virais do subtipo C (79). Embora os resultados obtidos não tivessem a partida, sido antecipados, na verdade a inespecificidade das amplificações obtidas reforçam a concepção original do método, o qual foi desenvolvido apenas para a identificação inequívoca de vírus dos subtipos B, G e formas genéticas CRF02_AG e CRF14_BG.

Considerações finais e perspectivas futuras

Na sua maioria, os estudos de caracterização genética do HIV-1 realizados em Portugal foram efectuados com base na análise de informação que nos revela uma imagem parcial do genoma viral (30, 75, 76, 78). Este tipo de abordagem, de facto, mostra-se incapaz de fornecer a informação suficiente para permitir uma caracterização integral do genoma viral. Esta limitação deve-se, em muito, à actividade recombinogénica da transcriptase reversa viral, a qual está na base da identificação em circulação de inúmeras formas recombinantes que comportam, num mesmo genoma, informação genética de vários subtipos virais. Face a isto, a sequenciação do genoma completo surge como o único método que garante uma caracterização genética clara das estirpes virais circulantes. Contudo, por ser um método dispendioso, tecnicamente exigente e laborioso é por isto, impraticável quando se tem em vista a análise de um elevado número de amostras. No entanto, para as regiões do Globo onde a epidemia do HIV-1 é caracterizada pela circulação de um número restrito de estirpes virais geneticamente distintas, ou onde a cocirculação de várias é dominada por um número mais restrito de algumas delas (como é o caso da epidemia Portuguesa) a implementação de ensaios de tipo MHA surgem como alternativas à sequenciação de genomas virais completos.

De uma forma geral, os ensaios de MHA convencionais consistem numa estratégia de amplificação por meio de duas reacções consecutivas para várias regiões do genoma (tantas quanto possível) numa abordagem por PCR em tempo-real. Os produtos amplificados na primeira reacção de amplificação são utilizados como matriz para a reacção subsequente, a qual é realizada na presença de sondas específicas para os subtipos/CRF a identificar, juntamente com um conjunto de *primers* de amplificação universal. Ao longo deste trabalho foram propostas estratégias experimentais distintas das que descrevem os ensaios de MHA convencionais, sendo que a abordagem mais aproximada à sua formulação convencional foi efectuada apenas para a região Gp41. No seu fundamento estas tiveram como propósito constituir uma alternativa passível de ser implementada num contexto financeiramente limitado, dado o enorme investimento requerido para a síntese de sondas de hidrólise específicas. Para além disto, apresentavam-se ainda, como uma mais-valia na análise de amostras em que uma baixa carga viral poderia dificultar a caracterização genética das estirpes virais nela presentes.

Em suma, foi possível desenvolver um conjunto de estratégias para a caracterização genética do HIV-1 em Portugal. Na bibliografia disponível sobre os ensaios de MHA os valores elevados de especificidade e sensibilidade a eles associados variam, normalmente, entre os 90 a 100% (46, 56, 57). Devido a limitações de tempo impostas não foi possível, à data da escrita deste manuscrito, confirmar por sequenciação e análise filogenética, todos os resultados obtidos de amplificação específica, produto das abordagens desenvolvidas no âmbito deste trabalho, não sendo, por consequência, possível quantificar de forma precisa a sensibilidade e especificidade das metodologias aqui apresentadas. De modo a avaliar correctamente a eficácia deste ensaio de MHA é de extrema prioridade a análise filogenética de todas as amostras clínicas incluídas neste estudo. Para além disto, e devido ao sucesso conhecido das estratégias por PCR em tempo-real utilizando sondas de hidrólise, mantém-se a possibilidade de re-optimização dos ensaios para a região de GP41, bem como a sua extensão a outras regiões do genoma.

Tal como descrito na bibliografia são esperadas ligeiras variações em função da região analisada (119). De facto, a experiência adquirida ao longo deste trabalho permitiu-nos detectar regiões ao longo do genoma cuja caracterização representa desafios acrescidos ao desenvolvimento deste tipo de abordagem. Apesar disto, o

cruzamento de resultados obtidos permitiu considerar as estratégias aqui propostas como uma ferramenta útil para o rastreio de elevados números de amostras, num curto espaço de tempo. Um utilizador devidamente treinado deverá conseguir analisar entre a 80 a 100 amostras por semana. Contudo, na eventualidade de surgirem resultados contraditórios, como exemplo, amplificações específicas com todos os *primers*, poderá ser necessária uma abordagem mais robusta com a finalidade de esclarecer a situação. Para tal, e alternativamente à sequenciação do genoma completo, poderá ser utilizado a estratégia aqui proposta de sequenciar fragmentos que incluam as regiões em estudo ao longo do genoma.

5. Referências bibliográficas

5. Referências bibliográficas

1. **Abd-Elsalam, A. K.** 2003. Bioinformatic tools and guideline for PCR primer design. *African Journal of Biotechnology* **2**:91-95.
2. **Abecasis, A.B., P. Lemey, N. Vidal, T. De Oliveira, M. Peeters, R. Camacho, B. Shapiro, A. Rambaut, and A. M. Vandamme.** 2007. Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form. *J. Virol.* **81**: 8543-8551.
3. **Aggarwal, I., M. Smith, I. D. Tatt, S. Murad, N. Osner, A. M. Geretti, and P. J. Easterbrook.** 2006. Evidence for onward transmission of HIV-1 non-B subtype strains in the United Kingdom. *J Acquir Immune Defic Syndr.* **41**:201-9.
4. **An, W., and A. Telesnitsky.** 2002. Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J. Virol.* **76**:7897-7902.
5. **Applied Biosystems.** Real-Time PCR Vs. Traditional PCR. pp 1-15. (disponível em <http://www.appliedbiosystems.com>).
6. **Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**:868-871.
7. **Buonaguro, L., M. Tagliamonte, M. L. Tornesello, and F. M. Buonaguro.** 2005. Evaluation of a modified version of Heteroduplex Mobility Assay for rapid screening of HIV-1 isolates in epidemics characterized by mono/dual clade predominance. *J Virol Methods.* **124**:123-34.

8. **Buonaguro, L., M. L. Tornesello, and F. M. Buonaguro.** 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* **81**:10209-10219.
9. **Butler, I. F., I. Pandrea, P. A. Marx, and C. Apetrei.** 2007. HIV Genetic diversity: Biological and public health consequences. *Curr. HIV Res.* **5**:23-45.
10. **Cane, P. A.** 2009. New developments in HIV drug resistance. *J. Antimicrob. Chemother.* **64**: 37-40.
11. **Ceccherini-Silberstein, F., I. Malet, R. D'Arrigo, A. Antinori, A. G. Marcelin, and C. F. Perno.** 2009. Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev.* **11**:17-29.
12. **Centers for Disease Control and Prevention (CDC).** 1981. *Pneumocystis Pneumonia*-Los Angeles. Epidemiologic Notes and Reports. *Morb. Mortal. Wkly. Rep.* **30**:1-3.
13. **Centlivre, M., M. Sala, S. Wain-Hobson, and B. Berkhout.** 2007. In HIV-1 pathogenesis the die is cast during primary infection. *AIDS* **21**:1-11.
14. **Centro de Vigilância Epidemiológica das Doenças Transmissíveis, INSA.** 2009. Infecção VIH/SIDA: A Situação em Portugal.
15. **Cerqueira, D. M., E. D. Ramalho, C. P. Oliveira, R. R. Silva, M. Franchini, M. S. Felipe, and C. R. Martins.** 2004. HIV-1 subtypes and mutations associated to antiretroviral drug resistance in human isolates from Central Brazil. *Braz.J. Microbiol.* **35**:187-192.
16. **Chory, J., and J. D. Jr. Pollard.** 2001. Separation of small DNA fragments by conventional gel electrophoresis. *Curr Protoc Mol Biol.* **2**:Unit2.7.

17. **Chung, C. T., S. L. Niemela, and R. H. Miller.** 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U.S.A.* **86**:2172-2175.

18. **Clapham, P. R., and A. McKnight.** 2002. Cell surface receptors, virus entry and tropism of primate lentiviruses. *J. Gen. Virol.* **83**:1809-1829.

19. **Clavel, F., D. Guetard, F. Brun-Vezinet, S. Chamaret, M. A. Rey, M. O. Santos-Ferreira, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, D. Klatzmann, J. L. Champalimaud, and L. Montagnier.** 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**:343-346.

20. **Coffin, J. M.** 1999. Molecular Biology of HIV, p. 3-40. *In* K. A. Crandall (ed.), *The Evolution of HIV*. The Johns Hopkins University Press, London.

21. **Cohen, M. S., N. Hellmann, J. A. Levy, K. DeCock, and J. Lange.** 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *J Clin Invest.* **118**:1244-54.

22. **De Baar, M. P., E. C. Timmermans, M. Bakker, E. de Rooij, B. van Gemen, and J. Goudsmit.** 2001. One-tube real-time isothermal amplification assay to identify and distinguish human immunodeficiency virus type 1 subtypes A, B, and C and circulating recombinant forms AE and AG. *J Clin Microbiol.* **39**:1895-902.

23. **Delgado, E., M. M. Thomson, M. L. Villahermosa, M. Sierra, A. Ocampo, C. Miralles, R. Rodríguez-Pérez, J. Diz-Aren, R. Ojea-de Castro, E. Losada, M. T. Cuevas, E. Vázquez-de Parga, R. Carmona, L. Pérez-Álvarez, L. Medrano, L. Cuevas, J. A. Taboada, and R. Nájera.** 2002. Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* **29**:536-43.

24. **Delwart, E. L., B. Herring, A. G. Rodrigo and J. I. Mullins.** 1995. Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. *PCR Methods Appl.* **4**:S202-16.
25. **Delwart, E. L., E. G. Shpaer, J. Louwagie, F. E. McCutchan, M. Grez, H. Rübsamen-Waigmann, and J. I. Mullins.** 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. *Science.* **262**:1257-61.
26. **Delwart, E. L., H. W. Sheppard, B. D. Walker, J. Goudsmit, and J. I. Mullins.** 1994. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. *J Virol.* **68**:6672-83.
27. **Dieffenbach, C. W., T. M. Lowe, and G. S. Dveksler.** 1993. General concepts for PCR primer design. *PCR Methods Appl.* **3**:S30-7.
28. **Erlich, H. A., D. Gelfand, and J. J. Sninsky.** 1991. Recent advances in the polymerase chain reaction. *Science.* **252**:1643-51.
29. **Esteves, A. R. Parreira, J. Piedade, T. Venenno, M. Franco, J. Germano De Sousa, L. Patrício, P. Brum, A. Costa, and W. F. Canas-Ferreira.** 2003. Spreading of HIV-1 Subtype G and *envB/gagG* Recombinant Strains among Injecting Drug Users in Lisbon, Portugal. *AIDS Res. Hum. Retroviruses* **19**:511-517.
30. **Esteves, A. R., Parreira, T. Venenno, M. Franco, J. Piedade, J. Germano De Sousa, and W. F. Canas-Ferreira.** 2002. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes *AIDS Res. Hum. Retroviruses* **18**:313-325.
31. **Felsenstein, J.** 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**:783-791.

32. **Flint, S. J., L. W. Enquist, R. M. Krug, V. R. Racaniello, and A. M. Skalka.** 2000. Principles of Virology: Molecular Biology, Pathogenesis, and Control. American Society for Microbiology, Washington, D.C.
33. **Forsman, A., and R. A. Weiss.** 2008. Why is HIV a pathogen? Trends Microbiol. **16**:555-560.
34. **Frankel, A.D., and J. A. Young.** 1998. HIV-1: fifteen proteins and an RNA. Annu. Rev. Biochem. **67**:1-25.
35. **Freed, E. O.** 2002. HIV-1 replication. Somat. Cell Mol. Genet. **6**:13-33.
36. **Gao, F. C., Yalu, N. L. David, J. A. Conway, T. B. Kepler, and H. Hui.** 2004. Unselected Mutations in the Human Immunodeficiency Virus Type 1 Genome Are Mostly Nonsynonymous and Often Deleterious. Journal of Virology, **78**:2426-2433.
37. **Gao, F., E. Bailes, D. L. Robertson, Y. Chen, M. C. Rodenburg, S. F. Michael, L.B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature **397**:436-441.
38. **Gendelman, H. E., T. S. Theodore, R. Willey, J. McCoy A. Adachi, R.J. Mervis, S. Venkatesan, and M. A. Martin.** 1987. Molecular characterization of a polymerase mutant human immunodeficiency virus. Virology **160**: 323-329.
39. **Gottlieb, M. S, R. Schroff, H. M. Schanker, J. D. Weisman, P, T, Fan, R. A. Wolf, and A. Saxon.** 1981. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men. N. Engl. J. Med. **305**:1425-1431.

40. **Granich, R., S. Crowley, M. Vitoria, Y. R. Lo, Y. Souteyrand, C. Dye, C. Gilks, T. Guerma, K. M. De Cock, and B. Williams.** 2010. Highly active antiretroviral treatment for the prevention of HIV transmission. *J Int AIDS Soc.* **13**:1.
41. **Gunson, R.N., T.C. Collins, and W.F. Carman.** 2006. Practical experience of high throughput real time PCR in the routine diagnostic virology setting. Review. *Journal of Clinical Virology*, **35**: 355–367.
42. **Hahn, B. H., G. M. Shaw, K. M. De Cock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. *Science* **287**: 607-614.
43. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**: 95-98.
44. **Harwood, A. J.** 1996. Basic DNA and RNA protocols. Introduction. *Methods Mol Biol.* **58**:v-vi.
45. **Heeney, J. L., A.G. Dalgleish, and R.A. Weiss.** 2006. Origins of HIV and the evolution of resistance to AIDS. *Science* **313**: 462-466.
46. **Hierholzer, J., S. Montano, M. Hoelscher, M. Negrete, M. Hierholzer, M. M. Avila, M. G. Carrillo, J. C. Russi, J. Vinales, A. Alava, M. E. Acosta, A. Gianella, R. Andrade, J. L. Sanchez, G. Carrion, J. L. Sanchez, K. Russell, M. Robb, D. Birx, F. McCutchan, and J. K. Carr.** 2002. Molecular Epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. *AIDS Res Hum - 50. Retroviruses.***18**:1339.
47. **Hoelscher, M., W. E. Dowling, E. Sanders-Buell, J. K. Carr, M. E. Harris, A. Thomschke, M. L. Robb, D. L. Birx, and F. E. McCutchan.** 2002. Detection of HIV-1 subtypes, recombinants, and dual infections in east Africa by a multi-region hybridization assay. *AIDS.* **16**:2055-64.

48. **Hraber, P., C. Kuiken, M. Waugh, S. Geer, W. J. Bruno, and T. Leitner.** 2008. Classification of hepatitis C virus and human immunodeficiency virus-1 sequences with the branching index. *J Gen Virol.* **89**:2098-107.
49. **Jetzt, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol.* **74**:1234-40.
50. **Kämpke, T.** 2007. PCR Primer Design, p75-91. *In* Yuryev, A. (ed) *Methods in Molecular Biology*, Volume 402, Humana Press, Totowa, NJ.
51. **Kandathil, A. J., S. Ramalingam, R. Kannangai, S. David, and G. Sridharan.** 2005. Molecular epidemiology of HIV. *Indian J. Med. Res.* **121**:333-344.
52. **Kato, S., R. Saito, Y. Hiraishi, N. Kitamura, T. Matsumoto, H. Hanabusa, M. Kamakura, Y. Ikeda, and M. Negishi.** 2003. Differential prevalence of HIV type 1 subtype B and CRF01_AE among different sexual transmission groups in Tokyo, Japan, as revealed by subtype-specific PCR. *AIDS Res Hum Retroviruses.* **19**:1057-63.
53. **Katoh, K., and H. Toh.** 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* **9**:286-98.
54. **Kaushik, S., M. Vajpayee, N. Wig, and P. Seth.** 2005. Characterization of HIV-1 Gag-specific T cell responses in chronically infected Indian population. *Clin Exp Immunol.* **142**:388-97.
55. **Keele, B. F., F. V. Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Y. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. H. Hahn.** 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**: 523-526.

56. **Kijak, G. H., E. Sanders-Buell, N. D. Wolfe, E. Mpoudi-Ngole, B. Kim, B. Brown, M. L. Robb, D. L. Birx, D. S. Burke, J. K. Carr, and F. E. McCutchan.** 2004. Development and application of a high-throughput HIV type 1 genotyping assay to identify CRF02_AG in West/West Central Africa. *AIDS Res Hum Retroviruses*. **20**:521-30.

57. **Kijak, G. H., S. Tovanabutra, E. Sanders-Buell, V. Watanaveeradej, M. S. de Souza, K. E. Nelson, V. Ketsararat, V. Gulgolarn, M. Wera-arpachai, S. Sriplienchan, C. Khamboonrueng, D. L. Birx, M. L. Robb, and F. E. McCutchan.** 2007. Distinguishing molecular forms of HIV-1 in Asia with a high-throughput, fluorescent genotyping assay, MHAbce v.2. *Virology*. **358**:178-91.

58. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. **16**:111-120.

59. **Knipe, D. M., P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus.** 2001. *Fields Virology*. Fourth edition. Lippincott Williams & Wilkins, Philadelphia, USA.

60. **Kosakovsky, P. S. L, and D. M. Smith.** 2009. Are all subtypes created equal? The effectiveness of antiretroviral therapy against non-subtype B HIV-1. *Clin Infect Dis*. **48**:1306-9.

61. **Kubista, M., J. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A Ståhlberg, and N. Zoric.** 2006. The real-time polymerase chain reaction. Review. *Molecular Aspects of Medicine*, **27**:95-125.

62. **Lal, R. B., S. Chakrabarti, and C. Yang.** 2005. Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development. *Indian J. Med. Res.* **121**: 287-314.
63. **Levy, D. N., G. M. Aldrovandi, O. Kutsch, and G. M. Shaw.** 2004. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci.* **101**:4204-4209.
64. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152-160.
65. **Mackay, I. M., K. E. Arden, and A. Nitsche.** 2002. Real-time PCR in virology. *Nucleic Acids Res.* **30**:1292-305.
66. **Mangeat, B., G. Gers-Huber, M. Lehmann, M. Zufferey, J. Luban, and V. Piguet.** 2009. HIV-1 Vpu neutralizes the antiviral factor Tetherin/BST-2 by binding it and directing its beta-TrCP2-dependent degradation. *PLoS Pathog.* **5**:e1000574.
67. **Marx, P. A.** 2005. Unsolved Questions over the Origin of HIV and AIDS: Although HIV originated in nonhuman primates, we do not understand how it became a successful pathogen in its human host. *ASM News* **71**:15-20.
68. **Miyauchi, K, Y. Kim, O. Latinovic, V. Morozov, and G. B. Melikyan.** 2009. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* **137**:433-44.
69. **Mathers, B. M., L. Degenhardt, B. Phillips, L. Wiessing, M. Hickman, S. A. Strathdee, A. Wodak, S. Panda, M. Tyndall, A. Toufik, R. P. Mattick, and 2007 Reference Group to the UN on HIV and Injecting Drug Use.** 2008. *Global*

- epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *Lancet*. **372**:1733-45.
70. **Moyle, G., J. Gatell, C. F. Perno, W. Ratanasuwan, M. Schechter, and C. Tsoukas.** 2008. Potential for new antiretrovirals to address unmet needs in the management of HIV-1 infection. *AIDS Patient Care STDS*. **22**:459-471.
71. **Nadai, Y., L. M. Eyzaguirre, N. T. Constantine, A. M. Sill, F. Cleghorn, W. A. Blattner, and J. K. Carr.** 2008. Protocol for nearly full-length sequencing of HIV-1 RNA from plasma. *PLoS One*. **3**:e1420.
72. **Nelson, J. A., S. A. Fiscus, and R. Swanstrom.** 1997. Evolutionary variants of the human immunodeficiency virus type 1 V3 region characterized by using a heteroduplex tracking assay. *J Virol*. **71**:8750-8.
73. **Nisole, S., N., and A. Saïb.** 2004. Early steps of retrovirus replicative cycle. *Retrovirology* **1**:9.
74. **Osmanov, S., C. Pattou, N. Walker, B. Schwardländer, J. Esparza, and WHO-UNAIDS Network for HIV Isolation and Characterization.** 2002. Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr*. **29**:184-90.
75. **Pádua, E., R. Parreira, R. Tendeiro, B. Nunes, J. Castela, I. Soares, A. Mouzinho, E. Reis, and M. T. Paixão.** 2009. Potential impact of viral load and genetic makeup of HIV type 1 on mother-to-child transmission: characterization of *env*-C2V3C3 and *nef* sequences. *AIDS Res Hum Retroviruses*. **25**:1171-7.
76. **Palma, A.C., F. Araújo, V. Duque, F. Borges, M.T. Paixão, and R. Camacho.** 2007. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.* **7**:391-398.

77. **Pantaleo, G., C. Graziosi, and A. S. Fauci.** 1993. The immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med.* **328**:327-35.
78. **Parreira, R., E. Pádua, J. Piedade, T. Venenno, M. T. Paixão, and A. Esteves.** 2005. Genetic analysis of human immunodeficiency virus type 1 *nef* in Portugal: subtyping, identification of mosaic genes, and amino acid sequence variability. *J Med Virol.* **77**:8-16.
79. **Parreira, R., J. Piedade, A. Domingues, D. Lobão, M. Santos, T. Venenno, J. L. Baptista, S. A. Mussa, A. T. Barreto, A. J. Baptista, and A. Esteves.** 2006. Genetic characterization of human immunodeficiency virus type 1 from Beira, Mozambique. *Microbes Infect.* **8**:2442-51.
80. **Peeters, M.** 2000. Recombinant HIV sequences: Their role in the global epidemic, p. 39-54. *In* C. Kuiken, B. Foley, B. Hahn, P. Marx, F. Mccutchan, J.W. Mellors, S. Wolinsky and B. Korber (ed.), *HIV Sequence Compendium 2000*, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.
81. **Peeters, M., and P. M. Sharp.** 2000. Genetic diversity of HIV-1: the moving target. *AIDS*, **14**:S129-40.
82. **Peeters, M., C. Toure-Kane, and J. N. Nkengasong.** 2003. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS* **17**:2547-2560.
83. **Peterlin, B. M., and D. Trono.** 2003. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat. Rev. Immunol.* **3**:97-107.
84. **Picker, L. J.** 2006. Immunopathogenesis of acute AIDS virus infection. *Curr. Opin. Immunol.* **18**:399–405.

85. **Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemée, F. Damond, D. L. Robertson, and F. Simon.** 2009. A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15**:871-2.
86. **Popovic, M., M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**:497-500.
87. **Quinn, T. C.** 2008. HIV epidemiology and the effects of antiviral therapy on long-term consequences. *AIDS*. 22 Suppl **3**:S7-12.
88. **Ravanfar, P., N. Mendoza, A. Satyaprakash, and B. I. Jordan.** 2009. HIV vaccines under study. *Dermatol Ther.* **22**:158-67.
89. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky and B. Korber.** 2000. HIV-1 nomenclature proposal. *Science.* **288**:55-6.
90. **Roux, K. H.** 1995. Optimization and troubleshooting in PCR. *PCR Methods Appl.* **4**:S185-94.
91. **Rozanov, M., U. Plikat, C. Chappey, A. Kochergin, and T. Tatusova.** 2004. A web-based genotyping resource for viral sequences. *Nucleic Acids Res.* **32**:654-659.
92. **Rychlik, W., W. J. Spencer and R. E. Rhoads.** 1990. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **18**:6409-12.

93. **Saitou, N., and N. Masatoshi.** 1987. The Neighbour-Joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.
94. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
95. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A* **74**:5463-5467.
96. **Sharp, P. M., and Hahn, B. H.** 2008. AIDS: prehistory of HIV-1. *Nature.* **455**:605-6.
97. **Sheffield, V. C., D. R. Cox, L. S. Lerman, and R. M. Myers.** 1989. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci U S A.* **86**:232-6.
98. **Sierra, S., B. Kupfer, and R. Kaiser.** 2005. Basics of the virology of HIV-1 and its replication. *J. Clin. Virol.* **34**:233-244.
99. **Siliciano, R. F (ed.).** 2006. *Acquired immune deficiency syndrome (AIDS).* Wiley, J. Encyclopedia of Life Sciences. (disponível em <http://www3.interscience.wiley.com>).
100. **Singh, K. V., and A. Kumar.** 2001. PCR Primer Design. *Molecular Biology Today.* **2**: 27-32.
101. **Smith, S. M.** 2006. The pathogenesis of HIV infection: stupid may not be so dumb after all. *Retrovirology* **3**: 60.

102. **Spira, S., M. A. Wainberg, H. Loemba, D. Turner, and B. G. Brenner.** 2003. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J. Antimicrob. Chemother.* **51**:229-240.
103. **Stebbing, J., B. Gazzard, and D. C. Douek.** 2004. Where does HIV live? *N Engl J Med.* **350**:1872-80.
104. **Steegen, K., E. Demecheleer, N. De Cabooter, D. Nges, M. Temmerman, P. Ndumbe, K. Mandaliya, J. Plum, and C. Verhofstede.** 2006. A sensitive in-house RT-PCR genotyping system for combined detection of plasma HIV-1 and assessment of drug resistance. *J Virol Methods.* **133**:137-45.
105. **Stürmer, M., W. Preiser, P. Gute, G. Nisius, and H. W. Doerr.** 2004. Phylogenetic analysis of HIV-1 transmission: *pol* gene sequences are insufficient to clarify true relationships between patient isolates. *AIDS.* **18**:2109-13.
106. **Takebe, Y., S. Kusagawa, and K. Motomura.** 2004. Molecular epidemiology of HIV: Tracking AIDS pandemic. *Pediatr. Int.* **46**:236-244.
107. **Takebe, Y., R. Uenishi, and X. Li.** 2008. Global molecular epidemiology of HIV: understanding the genesis of AIDS pandemic. *Adv Pharmacol.* **56**:1-25.
108. **Tamura K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596-1599.
109. **Tan, J. X., X. J. Kang, W. Zhang, P. P. Liu, X. Tong, and R. G. Yang.** 2007. Full-length clone and characterization of a human immunodeficiency virus type 1 subtype B' isolated from Hubei Province, China. *Chin Med J.* **120**:831-3.
110. **Tebit, D. M., L. Zekeng, L. Kaptué, H. G. Kräusslich, and O. Herchenröder.** 2003. Construction and characterisation of a full-length infectious molecular clone

- from a fast replicating, X4-tropic HIV-1 CRF02_AG primary isolate. *Virology*. **313**:645-52.
111. **Turner, D. H.** 1966. Thermodynamics of base pairing. *Curr Opin Struct Biol*. **6**:299-304.
112. **Uchil, P. D., and W. Mothes.** 2009. HIV Entry Revisited. *Cell* **137**:402-404.
113. **UNAIDS.** 2008. Report on the global HIV/AIDS epidemic 2008. WHO Library Cataloguing-in-Publication Data.
114. **Upchurch, D. A., R. Shankarappa, and J. I. Mullins.** 2000. Position and degree of mismatches and the mobility DNA heteroduplexes. *Nucleic Acids Res.* **28**:E69.
115. **Van Pelt-Verkuil, E., A. Van Belkum, and P. J Hays.** 2008. Principles and Technical Aspects of PCR Amplification. Hardbound ed., Springer.
116. **Vergne, L., and M. Peeters.** 2007. The challenge of HIV sequence diversity in the envelope glycoprotein, p.17-29. *In* J. D. Reeves and C. A. Derdeyn (ed.), Entry inhibitors in HIV therapy, Birkhäuser Basel, Montpellier.
117. **Vidal, N., S. E. Bazepeo, C. Mulanga, E. Delaporte, and M. Peeters.** 2009. Genetic characterization of eight full-length HIV type 1 genomes from the Democratic Republic of Congo (DRC) reveal a new subsubtype, A5, in the A radiation that predominates in the recombinant structure of CRF26_A5U. *AIDS Res. Hum. Retroviruses.* **25**:823-832.
118. **Wainberg, M. A., and J. Kuan-Teh.** 2008. 25 years of HIV-1 research- progress and perspectives. *BMC Med.* **6**: 1-7.
119. **Watanaveeradej, V., M. W. Benenson, M. D. Souza, N. Sirisopana , S. Nitayaphan, C. Tontichaivanich, R. Amphaipit, P. O. Renzullo, A. E. Brown,**

- J. G. McNeil, M. L. Robb, D. L. Birx, S. Tovanabutra, J. K. Carr and F. E. McCutchan . 2006.** Molecular epidemiology of HIV Type 1 in preparation for a Phase III prime-boost vaccine trial in Thailand and a new approach to HIV Type 1 genotyping. *AIDS Res Hum Retroviruses*. **22**:801-7.
120. **Wilson, J. W., P. Bean, T. Robins, F. Graziano, and D. H. Persing.** 2000. Comparative evaluation of three human immunodeficiency virus genotyping systems: the HIV-GenotypR method, the HIV PRT GeneChip assay, and the HIV-1 RT line probe assay. *J Clin Microbiol*. **38**:3022-8.
121. **Wolfe, N. D., W. M. Switzer, J. K. Carr, V. B. Bhullar, V. Shanmugam, U. Tamoufe, A. T. Prosser, J. N. Torimiro, A. Wright, E. Mpoudi-Ngole, F. E. McCutchan, D. L. Birx, T. M. Folks, D. S. Burke, and W. Heneine.** 2004. Naturally acquired simian retrovirus infections in central African hunters. *Lancet*. **363**:932-937.
122. **Worobey, M., M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, and E. Van Marck.** 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**:661-664.

6. Anexos

Anexo 1

Resumo dos resultados obtidos através das amplificações específicas por PCR convencional e em tempo-real e da análise filogenética para as amostras biológicas utilizadas neste trabalho. Para as amostras que foi possível obter informação estão indicados, o sexo do indivíduo, a contagem de linfócitos T CD4⁺ e a carga viral e grupo de risco (**UDIs**), local de proveniência, **AS**- Hospital Professor Doutor Fernando da Fonseca (Amadora-Sintra); **TA**- Centro de Atendimento a Toxicodependentes das **Taipas** em Lisboa; **CS P.Ge**a - Centro de Saúde de **Ponta Ge**a, na cidade da Beira, província de Sofala, em Moçambique; **Agarose** – PCR em formato convencional (visualização em gel de agarose); **T-R** – PCR em tempo-real (*SyberGreen*); **CV** - Cabo-Verde; **P** - Portugal; **Ang** - República de Angola; **M** – Moçambicano; **neg** – origem racial negra, **cauc** – origem racial caucasiano; **A.Filo.** – Análise filogenética; Para as regiões e amostras disponíveis estão ainda assinalados as informações referentes à subtipagem obtida previamente por HMA (**HMA**) e análise filogenética (comunicação pessoal S. Castro -assinalados com *) a partir de trabalhos anteriores. As amostras analisadas filogeneticamente neste trabalho encontram-se representadas a negrito (**B** - subtipo B, **G** - subtipo G, **02** - CRF02_AG), sendo que as que apresentaram resultados discordantes com a amplificação específica encontram-se a sublinhado (como exemplo, **B**). **N** - Resultados de amplificações indiscriminadas, seguindo-se o resultado da identificação do subtipo (como exemplo, BG). As amostras analisadas por PCR em tempo-real com sondas de hidrólise para a região de Gp41 encontram-se assinaladas com o símbolo “+”, para as que apresentaram um resultado coincidente e detectável, estão ainda assinaladas as que não foram detectadas por esta abordagem (-). Nota: Estão apenas representados os resultados da análise filogenética obtidos até à data uma vez que esta análise encontrava-se em curso à data de execução deste relatório.

Amostra	Sexo	Idade	Origem	Nacionalidade	Grupo de Risco	UDIs	C/ CD4 (células/µl)	Carga Viral (cópias/ml)	Região analisada														
									Gag			Pr		RT		Rev		Gp120			Gp41		LTR
									HMA	Agarose	T-R	Agarose	T-R	Agarose	T-R	Agarose	T-R	HMA	Agarose	T-R	Agarose	T-R	A.Filo.
709	M	31	AS	CV/neg	x	167	59589		A	B	B	B	B	B	B	B	B	B	B	B	B+	-	
508	M	38	AS	P/cauc.	x	42	174012		G	G	G B	B	B	N (BG)	N (BG)	G	G	B	N (BG)	B	B	B-	-
988	F	23	AS	P/cauc.	x	446	12937		IBNG	02	02	G	G	02	02	G	G	A	N (B02G)	N (BG)	G	G-	-

317	M	33	AS	P/cauc.	x	14	106294	D	B	B D	B	B	B	B	B	B	D	B	B	B	B+	-
658	M	34	AS	P/cauc.	x	361	3685	G	G	G	G	G F1	G	G	G	G	G	G	G	G	G+	-
893	M	45	AS	P/cauc.	x	199	39847	B	B	B	B	B	B	B	B	B	B	B	B	B	B	-
727	M	34	AS	P/cauc.	x	46	284331	IBNG	02	02	G	G	02	02	G	G	A	N (B02G)	N (BG)	G	G+	-
423	M	30	AS	P/cauc.	x	622	29878	B	B	B	B	B	2	B	B	B	B	B	B	B	B+	-
227	F	36	AS	P/cauc.	x	619	49632	G	G	G	G	G	G	G	G	G	B	N (BG)	N (BG)	G	G+	-
3013	M	23	AS	P/cauc.	x	443	58008	B	B	B	B	B	B	B	B	N (BG)	B	B	B	B	B+	-
9000	M	26	AS	P/cauc.	x	737	5034	G	G	G	G	G	G	G	G	G	F	N (G02)	N (G02)	G	G+	-
919	M	29	AS	P/cauc.	x	278	50	G	G	G	G	G	0	0	0	B	B	B	B	0	G	-
6060	F	28	AS	P/cauc.	x	28	359405	G	G	G	G	G	G	G	N (BG)	G	B	B	B	G	G+	-
983	M	27	AS	P/cauc.	x	492	28326	B	B	B	G	G	B	B	N (BG)	B	B	N (BG)	B	B	B-	-
695	F	37	AS	E/cauc	x	513	33558	B	B	B	B	B	B	B	B	B	B	N (BG)	B	N (BG)	B+	-
806	F	20	AS	P/cauc.	x	1 021	26464	G	G	G	B	B	G	G	N (BG)	G	G	G	G	N (BG)	G-	-
251	M	30	AS	P/cauc.	x	727	16751	G	N (G02)	G	G	G	G	G	G	G	G	G	G	G	G+	-
482	F	23	AS	P/cauc.	x	140	3714	G	G	G	N (BG)	G	G	G	G	G	B	B	B	G	G-	-
271	F	26	AS	P/cauc.	x	779	8002	G	G	G	N (BG)	G	G	G	G	G	G	N (BG)	G	0	0-	-
73	F	27	AS	P/cauc.	x	367	8593	B	N (B02)	B	B	B	B	B	B	B	B	B	B	B	B+	-
9108	F	22	AS	P/cauc.	x	344	811	G	G	G	N (BG)	G	G	G	G	G	B	G	G	G	G-	-
8731	M	21	AS	P/cauc.	x	302	6892	B	B	B	B	B	0	B	B	B	B	G	G	B	B+	-
1147	M	40	AS	P/cauc.	x	194	30894	B	B	B	B	B	B	B	B	B	B	B	B	B	B-	-
471	M	25	AS	P/cauc.	x	885	2748	-	N (02G)	N (02 G)	N (BG)	G	N (G02)	N (G02)	G	G	A	B	B	G	G-	-
1631	M	24	AS	P/cauc.	x	420	6270	-	G	G	G	G	G	G	G	G	G	G	G	G	G-	-
7139	F	26	AS	P/cauc.	x	323	7485	-	G	G	G	G	G	G	G	G	G	G	G	G	G-	-
1544	M	35	AS	P/cauc.	x	518	20344	-	G	G	G	G	G	G	G	G	G	G	G	G	G+	-
176	M	40	AS	P/cauc.	x	4312	22416	-	B	BG	N (BG)	G	B	B	G	G	G	G	G	0	G-	-
9456	M	32	AS	P/cauc.	x	567	4607	-	G	G	N (BG)	G	G	G	G	G	G	G	G	G	G	-
4132	M		AS	P/cauc.	x	484	2687		N (02B)	02	G	G	02	02	G	G	A	N (B02G)	N (G02)	G	G	-
4773	M	23	TA	P/cauc.	x	N/A	N/A	IBNG	02	02	G	G	02	02	G	G	A	N (G02)	N (G02)	G	G	-
836	M	27	TA	P/cauc.	x	N/A	N/A	-	G	G	G	G	02	02	G	G	A	02	02	G	G	-
5236	M	35	TA	P/cauc.	x	N/A	N/A	-	G	G	G	G	G	G	G	G	G	G	G	G	G-	-

955	M	27	TA	P/cauc.	x	N/A	N/A	-	G	G	G	G	G	G	G	G	F	B	B	G	G	-
2	F	42	TA	P	x	N/A	N/A	-	B	B	B	B*	B	B*	B	B	-	B	B	B	B	-
3	M	39	TA	P	x	N/A	N/A	-	G	G	N (BG)	G*	G	G*	G	G	-	G	G	G	G	-
13	M	40	TA	P	x	N/A	N/A	-	G	G	G	G*	N (G02)	N (G*02)	G	G	-	0	N (BG)	G	G	-
16	M	38	TA	P	x	N/A	N/A	-	B	B	B	B*	B	B*	B	B	-	B	B	B	B	-
24	M	38	TA	P	x	N/A	N/A	-	02	G	B	B*	B	B*	B	B	-	B	B	B	B	-
38	M	36	TA	P	x	N/A	N/A	-	02	02	B	B*	G	G B*	B	B	-	B	B	B	B	-
72	F	44	TA	P	x	N/A	N/A	-	G	G	N (BG)	G*	G	G*	N (BG)	N (BG)	-	G	G	0	B	-
85	M	32	TA	Ang	x	N/A	N/A	-	G	G	N (BG)	G	G	G*	G	G	-	B	B	G	G	-
98	M	29	TA	P	x	N/A	N/A	-	G	G	G	G*	G	G*	G	G	-	G	G	G	G	-
107	F	31	TA	P	x	N/A	N/A	-	N (02G)	N (02 G)	G	G* <u>B</u>	B	B*	B	B	-	B	B	B	B	-
114	M	38	TA	P	x	N/A	N/A	-	G	G	B	B	B	B	N (B02)	B	-	B	B	B	B	-
PG81	F	21	CS P.Gea	M	N/A	N/A	N/A	C	N (B02G)	N (B02G)	N (BG)	N (BG)	N (B02)	N (G02)	G	G	C	N (B02G)	N (B02G)	0	0	C
PG176	F	21	CS P.Gea	M	N/A	N/A	N/A	C	N (B02G)	N (B02G)	N (BG)	N (BG)	0	2	0	0	C	N (B02G)	N (B02G)	0	0	C
PG178	F	20	CS P.Gea	M	N/A	N/A	N/A	C	N (B02G)	N (B02G)	N (BG)	N (BG)	N (G02)	N (G02)	0	0	C	N (B02G)	N (B02G)	0	0	C
PG184	F	31	CS P.Gea	M	N/A	N/A	N/A	C	N (B02G)	N (B02G)	N (BG)	N (BG)	N (G02)	2	0	0	C	N (B02G)	N (B02G)	0	0	C
PG187	F	45	CS P.Gea	M	N/A	N/A	N/A	C	N (B02G)	N (B02G)	N (BG)	N (BG)	N (BG)	2	0	0	C	N (B02G)	N (B02G)	0	0	C

Anexo 2

Na tabela estão representados os *Primers/sondas* utilizados nas amplificações de DNA efectuadas ao longo deste trabalho por PCR e PCR em tempo-real. Estão representadas a posição relativa dos *primer* em relação ao HXB2, a orientação, a sequência nucleotídica e a temperatura de desnaturação (**TM**, inglês **Temperature Metlting**); + Amplificação universal; * Amplificação específica; ** Amplificação específica - sondas *TaqMan*; Gp12014B-3* A para não emparelhar com CRF02_AG; Gp120G-3* T para não emparelhar com CRF02_AG; G para não emparelhar com B e CFR02_AG; Int02G-2* T para não emparelhar com B.

Região	Nome do <i>Primer</i>		Posição HXB2*	Sequência do <i>primer</i> (5'---3')	TM (°C)	Directo/ Reverso
gag	H1Gag1584+	HMA (NIH)	1577-1595	AAAGATGGATAATCCTGGG	52.6	Directo
	g17+	HMA (NIH)	2017-2040	TCCACATTTCACAGCCCTTTT	61.0	Reverso
	GAGOF+	Este trabalho	1735-1759	TGGATGACAGAMACCTTGYTGGTCC	63.6 - 67.5	Directo
	Prot OR +	Este trabalho	2328-2355	CTTCTAAYACTGTATCATCTGCTCCTGT	57.7 - 60.3	Reverso
	GagIF+	Este trabalho	1817-1842	TAGAAGAAATGATGACAGCATGYCAG	57.2 - 59.6	Directo
	GagIR+	Este trabalho	2023-2048	CCYTCCTTYCCACATTTCACAGCC	64.8 - 69.4	Reverso
	GagB*	Este trabalho	1878-1905	GGCTGAAGCAATGAGCCAAGTAACAAAT	61.9	Directo
	Gag02A*	Este trabalho	1878-1905	GGCYGAGGCAATGAGTCAAGYACAACAR	63.7 - 68.8	Directo
	Gag14G*	Este trabalho	1878-1905	AGCTGAGGCAATGAGCCWGGCATCAGGK	70.6 - 72.1	Directo
pol (Pr)	PrRtOF+	Este trabalho	2328-2355	ACAGGAGCAGATGATACAGTRTTAGAAG	57.7 - 60.3	Directo
	PrRtOR+	Este trabalho	2979-3010	TTCCATCCYTGTTGAAGCACATTGTACTGATA	63.6 - 65.4	Reverso
	PrRtIF+	Este trabalho	2375-2404	ATGGAARCCAAAAATGATAGGGGGAATTGG	63.2 - 64.8	Directo

	PrRtIR+	Este trabalho	2581-2606	TGGRCCATCCATTCCTGGCTTTAATT	62.1 - 64.5	Reverso
	PrRtB*	Este trabalho	2433-2462	CARATACYCATAGAAATCTGYGGACATAAA	55.2 - 60.7	Directo
	PrRt0214G*	Este trabalho	2433-2462	CARATACTTATAGAAATTTGTGGAAAAARG	51.8 - 54.8	Directo
Pol (RT)						
	RT1+	Ref.235	2604-2632	CCAAAAGTTAAACAATGGCCATTGACAGA	60.8	Directo
	RT4+	Ref.235(adaptado)	3231-3253	TGGAGTTCATAWCCCATCCAAAG	58.5 - 58.9	Reverso
	RTOF+	Este trabalho	2982-3008	CAGTACAATGTGCTTCCACARGGATGG	63.2 - 65.3	Directo
	RTOR+	Este trabalho	3627-3653	TAAYTGYTTTACATCATTAGTGTGRGC	53.2 - 59.1	Reverso
	RTIF+	Este trabalho	3301-3327	GCTGGACTGTCAAYGAYATACARAART	55.4 - 63.1	Directo
	RTIR+	Este trabalho	3554-3580	TCTTGATARATTTGRTATGTCCAYTGG	53.9 - 59.5	Reverso
	RTB*	Este trabalho	3422-3452	AGTAATACCACTAACAGAAGAAGCAGAGCTA	60.1	Directo
	RT02A*	Este trabalho	3422-3452	TATAGTAMCACTGACTGAGGAAGCAGAATTA	58.6 - 60.5	Directo
	RT14G*	Este trabalho	3422-3452	CATAGTAYCACTRACWGCAGAAGCAGAATTG	58.8 - 62.9	Directo
Pol (INT)						
	IntOF+	Este trabalho	4386-4411	CAAGTAGACTGTAGTCCAGGAATATG	57.3	Directo
	IntOR+	Este trabalho	4953-4983	GTATTACTACYGCCCCTTACCTTTCCAGAG	65.2 - 67.3	Reverso
	IntIF+	Este trabalho	4467-4494	GCCAGTGGMTAYATAGAAGCAGAAGTTA	57.9 - 62.3	Directo
	IntIR-1+	Este trabalho	4650-4679	TACTCCTTGRCCTTGGGGATTGTAGGGAAT	64.7 - 66.8	Reverso
	IntIR-2+	Este trabalho	4899-4924	TCTCTGCTGTCCCTGWAATAAACCCG	64.0 - 64.7	Reverso
	IntB-1*	Este trabalho	4613-4637	CGCCTGTTGGTGGGCRGGRATCAAG	73.1 - 77.3	Directo
	IntB-2*	Este trabalho	4712-4733	CTGATCTCTTACCTGTCCTATA	53.1	Reverso

	Int02G-1*	Este trabalho	4613-4641	AGCATGTTGGTGGGCAAATRTCACACARG	63.8 - 68.1	Directo
	Int02G-2*	Este trabalho	4712-4730	ATCCCTRACCTTCCIGATG	55.2 - 58.8	Reverso
Rev/gp120						
	VifOF+	Este trabalho	5558-5584	GATGGAACAAGCCCCAGARGACCARGG	68.5 - 73.2	Directo
	Gp120 OR+	Este trabalho	6556-6582	CACATGGYTTTAGGCTTTSRTCCATA	60.3 - 64.4	Reverso
	VpuIR+	Este trabalho	6203-6227	CATTKCCACTRTCTTCTGCTCTTTC	56.5 - 61.3	Reverso
	RevB*	Este trabalho	6035-6060	CTATCAAAGCAGTAAGTAGTAYATGT	50.7 - 53.4	Directo
	Rev0214G*	Este trabalho	6035-6053	GTACCAAAGCAGTRAGTARTATAAATTAR	50.6 - 54.8	Directo
	Gp120IF-1+	Este trabalho	6328-6356	GGGTACRGCTCTAYTATGGRTACCTGTG	63.5 - 70.0	Directo
	VprIF-2+	Este trabalho	5956-5984	CTTAGGCATCTCCTATGGCAGGAAGAAGC	65.7	Directo
	Gp120IR+	Este trabalho	6440-6466	GGGYTRGGTCTGTGGGTACACAGGCA	70.8 - 75.6	Reverso
	Gp12014B-1*	Este trabalho	6408-6438	GAHACAGAGGTACATAATGTTTGGGCCACAC	77,3	Directo
	Gp12014B-2*	Este trabalho	6358-6389	TGCACAAAATAGAGTGGTRGTTGCTTCYTTC	62.7 - 66.9	Reverso
	Gp12014B-3*	Este trabalho	6360-6379	AGAGTGGTGGTTGCTTCATT	57.3	Reverso
	Gp12002-1*	Este trabalho	6408-6438	GATRCAGAAGTACATAATGTCTGGGCTACAC	71,6	Directo
	Gp12002-2*	Este trabalho	6310-6340	TAGACCGTGACCCACAAMTYTTCAGCMT	63.2 - 68.6	Reverso
	Gp12002-3*	Este trabalho	6316-6335	CGTGACCCACAAMTYTTCAG	55.8 - 60.0	Reverso
	Gp120G-1*	Este trabalho	6408-6438	AGTWCTGAAARMCATAATGTCTGGGCTACAC	70,5	Directo
	Gp120G-2*	Este trabalho	6408-6438	GTGTAGCCCAGACATTATGGYTTTCAGTACT	62.2 - 63.7	Reverso
	Gp120G-3*	Este trabalho	6414-6438	CCCATACATTATGGYTGTTCAGTACT	56.9 - 58.8	Reverso
	ED31+	HMA (NIH)	6817-6845	CCTCAGCCATTACACAGGCCTGTCCAAAG	68.7	Directo

	ED33+	HMA (NIH)	7360-7381	TTACAGTAGAAAAATTCCCCTC	53.3	Reverso
gp41						
	gp41 ectO_for+	Andreia 2006	7642-7668	GAGATATGAGGGACAATTGGAGAAGTG	60.7	Directo
	gp41 ectO_rev+	Andreia 2006	8510-8533	CGGTGGTAGCTGAAGAGGCACAGG	68.2	Reverso
	Gp41OF+	Este trabalho	8340-8365	AATAGAGTTAGGMAGGGATAYTCACC	56.8 - 60.8	Directo
	Nef OR+	Este trabalho	9163-9189	TGGCCCTGGTGTGTARTTCTGCCAATC	66.8 - 68.6	Reverso
	Gp41IF+	Este trabalho	8511-8537	CTGTGYCTCTTCAGCTACCACCGMTG	64.4 - 68.6	Directo
	Gp41IR+	Este trabalho	8682-8707	CTRTCTGTCCMVTYAGCTACTRCTAT	52.4 - 65.0	Reverso
	Gp41B*	Este trabalho	8574-8594	CTTCTGGGACGCAGGGGGTGG	72	Directo
	Gp410214G*	Este trabalho	8587-8587	CAGCAGYCTCAAGGGACTGAGACT	64.4 - 67	Directo
	Gp41ProbeB**	Este trabalho	8574-8594	JOECTTCTGGGACGCAGGGGGTGGBHQ-1	72	Directo
	Gp41Probe0214G**	Este trabalho	8587-8587	RoxCAGCAGYCTCAAGGGACTGAGACTBHQ-2	64.4 - 67	Directo
nef						
	NefO1+	Ref. 212	8513-8533	GTGCCTCTTCAGCTACCACCG	64.3	Directo
	NefO2+	Ref.212	9488-9508	AGCATCTGAGGGTTAGCCACT	61.3	Reverso

Anexo 3

Development of a multi-region hybridization assay (MHA_{BG0214}) for the identification of subtypes B, G and recombinant forms CRF02_AG and CRF14_BG of the human immunodeficiency virus type 1 (HIV-1) circulating in Portugal

Ferdinando B. Freitas¹, Aida Esteves^{1,2}, João Piedade^{1,2}, and Ricardo Parreira^{1,2}

¹Unidade de Ensino e Investigação Virologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (UNL); ²Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (UNL).

Keywords: HIV-1; Multi-region hybridization assay; Molecular epidemiology; Portugal.

Since most methods used for the genetic characterization of HIV-1 are based on the analysis of specific regions, they only provide a fragmented view of the viral genome, and often fail to identify recombinant viruses. The most efficient method for HIV-1 genetic characterization is still full-genome sequencing, but the associated costs and low throughput preclude it from being routinely used for the analysis of large numbers of viral strains. In this context, the aim of our study was the development of a new multi-region hybridization assay (MHA_{BG0214}) for fast, relatively inexpensive, specific and high throughput genotyping of the major HIV-1 forms circulating in Portugal, where the epidemic, though characterized by a plethora of highly diverse strains, is dominated by subtypes B, G and CRF02_AG and CRF14_BG circulating recombinant viruses.

In its conventional real-time PCR format, MHA relies on the amplification by PCR of several regions (usually ≥ 5) scattered along the HIV-1 genome with so called universal primers. The obtained amplicons are then distributed to several second-round amplification reactions, each carried out with inner universal primers and a subtype-specific *TaqMan* probe. Based on full alignments of representative HIV-1 reference sequences, we designed universal and subtype-specific primers/probes for the amplification of 7 different regions of the viral genome: the Gag, Protease, Reverse Transcriptase, Integrase, Rev, Gp120 and Gp41 coding sequences. Optimization of reaction conditions, established using 7 HIV-1 references, then served as a starting point for the analysis of 50 HIV-1 strains circulating in Portugal, using either proviral DNA immobilized on FTA™ cards or viral RNA extracted from plasma samples. Universal and subtype-restrictive amplification conditions were set for 6 of the 7 targeted sections of the viral genome, the integrase coding region consistently revealing ambiguous subtyping results. MHA_{BG0214} was implemented using a real-time PCR-based approach, with amplicon detection dependent on the use of either *SYBR*® *Green* I (all regions) or a *TaqMan* probe (Gp41). Alternatively, a technically less demanding strategy based on conventional PCR and agarose gel analysis of reaction products was also developed. Regardless of the experimental schemes used, preliminary data disclosed high specificity of the obtained subtyping results for all regions of the HIV-1 genome analysed.